

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2003, 21:47:55 : Search time 935 Seconds
(without alignments)
108331.397 Million cell updates/sec

Title: US-10-001-857-42

Perfect score: 3096

Sequence: 1 ttcttcacgaactccagg.....atttcttgacaaaaaaa 3096

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2172232 seqs, 163554964 residues

Total number of hits satisfying chosen parameters: 4344464

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

```
1: /cgn2_6/ptodata/2/pubna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubna/US05_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubna/US09_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubna/US60_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|----------------------|
| 1 | 3096 | 100.0 | 3096 | 13 | US-10-001-857-42 |
| 2 | 2228.4 | 72.0 | 2610 | 12 | US-09-814-351-20725 |
| 3 | 560.8 | 18.1 | 745 | 12 | US-09-814-353-15927 |
| 4 | 509.4 | 16.5 | 610 | 12 | US-09-814-353-16155 |
| 5 | 496.2 | 16.0 | 602 | 12 | US-09-814-353-3218 |
| 6 | 496.2 | 16.0 | 602 | 12 | US-09-814-353-9543 |
| 7 | 451.2 | 14.6 | 482 | 11 | US-09-918-995-20914 |
| 8 | 434 | 14.0 | 434 | 12 | US-09-814-353-3455 |
| 9 | 434 | 14.0 | 434 | 12 | US-09-814-353-9771 |
| 10 | 350.8 | 11.3 | 398 | 10 | US-09-983-965-210 |
| 11 | 268.6 | 8.7 | 501 | 10 | US-09-983-590-5693 |
| 12 | 182 | 5.9 | 489 | 12 | US-10-027-632-323193 |
| 13 | 182 | 5.9 | 489 | 13 | US-10-027-632-323193 |
| 14 | 140.4 | 4.5 | 151 | 13 | US-10-001-857-41 |
| 15 | 87.8 | 2.8 | 815 | 13 | US-10-001-857-111 |
| 16 | 60 | 1.9 | 60 | 12 | US-09-908-975-16495 |

| | | | | | | |
|----|------|-----|--------|----|---------------------|--------------------|
| 17 | 58.2 | 1.9 | 3163 | 12 | US-10-017-161-1857 | Sequence 1857, Ap |
| 18 | 58 | 1.9 | 956 | 12 | US-10-027-632-31508 | Sequence 31508, A |
| 19 | 58 | 1.9 | 956 | 13 | US-10-027-632-31508 | Sequence 31508, A |
| 20 | 56.2 | 1.8 | 5452 | 12 | US-10-017-161-1481 | Sequence 1481, Ap |
| 21 | 56 | 1.8 | 1117 | 12 | US-10-017-161-1403 | Sequence 1403, Ap |
| 22 | 54.6 | 1.8 | 12733 | 14 | US-10-032-393-47 | Sequence 8, Appl |
| 23 | 54.6 | 1.8 | 12739 | 14 | US-10-032-393-8 | Sequence 8, Appl |
| 24 | 53.2 | 1.7 | 2561 | 10 | US-09-976-740-48 | Sequence 48, Appl |
| 25 | 53.2 | 1.7 | 2561 | 13 | US-10-023-529-48 | Sequence 48, Appl |
| 26 | 53.2 | 1.7 | 2561 | 13 | US-10-023-523-48 | Sequence 48, Appl |
| 27 | 52.8 | 1.7 | 3133 | 12 | US-10-017-161-1483 | Sequence 1483, Ap |
| 28 | 52.8 | 1.7 | 6432 | 12 | US-10-329-079-10 | Sequence 10, Appl |
| 29 | 52.8 | 1.7 | 37360 | 12 | US-10-329-079-6 | Sequence 6, Appl |
| 30 | 51.4 | 1.7 | 440 | 12 | US-10-063-685-52 | Sequence 52, Appl |
| 31 | 51.4 | 1.7 | 440 | 14 | US-10-184-644-202 | Sequence 202, Appl |
| 32 | 51.4 | 1.7 | 440 | 14 | US-10-184-634-202 | Sequence 202, Appl |
| 33 | 51 | 1.6 | 152331 | 13 | US-10-095-407-16 | Sequence 16, Appl |
| 34 | 50.6 | 1.6 | 65 | 12 | US-09-908-975-879 | Sequence 879, Ap |
| 35 | 50.6 | 1.6 | 922 | 12 | US-10-017-161-1435 | Sequence 1435, Ap |
| 36 | 50.4 | 1.6 | 883 | 12 | US-10-027-632-4358 | Sequence 4358, Ap |
| 37 | 50.4 | 1.6 | 883 | 13 | US-10-027-632-4358 | Sequence 4358, Ap |
| 38 | 50.2 | 1.6 | 1065 | 10 | US-09-804-682-33 | Sequence 33, Appl |
| 39 | 50 | 1.6 | 1064 | 10 | US-09-804-682-29 | Sequence 29, Appl |
| 40 | 50 | 1.6 | 1614 | 10 | US-09-976-740-45 | Sequence 45, Appl |
| 41 | 50 | 1.6 | 1614 | 13 | US-10-023-529-45 | Sequence 45, Appl |
| 42 | 50 | 1.6 | 1614 | 13 | US-10-023-523-45 | Sequence 45, Appl |
| 43 | 50 | 1.6 | 14225 | 10 | US-09-976-740-50 | Sequence 50, Appl |
| 44 | 50 | 1.6 | 14225 | 13 | US-10-023-529-50 | Sequence 50, Appl |
| 45 | 50 | 1.6 | 12425 | 13 | US-10-023-523-50 | Sequence 50, Appl |

ALIGNMENTS

RESULT 1

US-10-001-857-42
; Sequence 42, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P.
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 42
; LENGTH: 3096
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-857-42

| | | | | | | | |
|-----------------------|--------|------------------------------------------------------------|------|------------|----|--------|------|
| Query Match | 100.0% | Score | 3096 | DB | 13 | Length | 3096 |
| Best Local Similarity | 100.0% | Pred. No. | 0 | | | | |
| Matches | 3096 | Conservative | 0 | Mismatches | 0 | Indels | 0 |
| Gaps | 0 | | | | | | |
| Qy | 1 | TTCTTCACGAACCTCCAGGCGCTGTATAGGAACATATCCGTTGTCAGGACGAGTA | 60 | | | | |
| Db | 1 | TTCTTCACGAACCTCCAGGCGCTGTATAGGAACATATCCGTTGTCAGGACGAGTA | 60 | | | | |
| Qy | 61 | GCACGCTGTGTCTTCGAGGCTTGGCTCTGCTTCGTCGTCGCAACCACTAAGCTCTACG | 120 | | | | |
| Db | 61 | GCACGCTGTGTCTTCGAGGCTTGGCTCTGCTTCGTCGTCGCAACCACTAAGCTCTACG | 120 | | | | |
| Qy | 121 | CAAACTCCACGGTTTCTTCGCGCTTCGCTTCACCTTTCTAAGAAATTCACGAGGCA | 180 | | | | |
| Db | 121 | CAAACTCCACGGTTTCTTCGCGCTTCGCTTCACCTTTCTAAGAAATTCACGAGGCA | 180 | | | | |

181 GCGCAGACGGGCGGGCTCTGAGATCTCGGGCTCGCGCTCTTTTCGGGAAACCGCCACTA 240
181 GCGCAGACGGGCGGGCTCTGAGATCTCGGGCTCGCGCTCTTTTCGGGAAACCGCCACTA 240
241 CCCAGACTCCGACAGAGGGTGAAGAGATAAATTCGGGTCTCGCGATCGTCTTAATC 300
241 CCCAGACTCCGACAGAGGGTGAAGAGATAAATTCGGGTCTCGCGATCGTCTTAATC 300
301 TCGCAGAGAGAGAGCGGCGCGCCATCGGCCGAAACGAGGGCGGTGGCGAGGAGGGGGTG 360
301 TCGCAGAGAGAGAGAGCGGCGCGCCATCGGCCGAAACGAGGGCGGTGGCGAGGAGGGGGTG 360
361 TCGCCGGGAGCGCAAGTCCCGGAGTAAGGAGAGGGGCGGGTCCGCGTCCCG 420
361 TCGCCGGGAGCGCAAGTCCCGGAGTAAGGAGAGGGGCGGGTCCGCGTCCCG 420
421 GCATACGCAATGCTGACGCTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCG 480
421 GCATACGCAATGCTGACGCTCCCGTCCCGTCCCGTCCCGTCCCGTCCCG 480
481 GCGGAGGCGGCTGCTTATTCGTTGTCGACAGTGGCGGCGGGTGGCGGTCACACG 540
481 GCGGAGGCGGCTGCTTATTCGTTGTCGACAGTGGCGGCGGGTGGCGGTCACACG 540
541 GGAGAGTAGGCATATGTTGTAAGAGTCTCTGATGATGACGATTCAGGATGGAG 600
541 GGAGAGTAGGCATATGTTGTAAGAGTCTCTGATGATGACGATTCAGGATGGAG 600
601 CTCAGTATCCGAGAAAATGAGAAAAGCAATACAACTGGTGGTGGATACCAAGAT 660
601 CTCAGTATCCGAGAAAATGAGAAAAGCAATACAACTGGTGGTGGATACCAAGAT 660
661 TTGGAAGAGCTTGTGAGAAATGAGTGGAGAACTACTCATGATAAGCTATTTCGT 720
661 TTGGAAGAGCTTGTGAGAAATGAGTGGAGAACTACTCATGATAAGCTATTTCGT 720
721 CTTTGTGAAGCCATGCTCTATGAAATGATGATGATGATGATGATGATGATGAT 780
721 CTTTGTGAAGCCATGCTCTATGAAATGATGATGATGATGATGATGATGATGAT 780
781 GGAACCAAGTAAATCGAAGAGTCTCAATTTTGAACAGCTATCAAGATGGCACTATT 840
781 GGAACCAAGTAAATCGAAGAGTCTCAATTTTGAACAGCTATCAAGATGGCACTATT 840
841 AAAATTAAGATCTACCTTGCTGAACTGATAGGATATGATGATGATGATGATGAT 900
841 AAAATTAAGATCTACCTTGCTGAACTGATAGGATATGATGATGATGATGATGAT 900
901 TTGATAACGTTGTAAGGCGCAATTCACCTGCGACAGAGTATTTACGCGCTTACATT 960
901 TTGATAACGTTGTAAGGCGCAATTCACCTGCGACAGAGTATTTACGCGCTTACATT 960
961 CATATCCAGACTTTATAGAGATCTGCTATGAGGCTTTTGTCTGGGATCTTGAAA 1020
961 CATATCCAGACTTTATAGAGATCTGCTATGAGGCTTTTGTCTGGGATCTTGAAA 1020
1021 ATCTGTGATTCGAAGGAAAAGTAAATGAGTCTGTTTGAAGAGAGAGATTTT 1080
1021 ATCTGTGATTCGAAGGAAAAGTAAATGAGTCTGTTTGAAGAGAGAGATTTT 1080
1081 CAGTCAATGACTTATGGAATTTAAATGGCTAAACAGTGTGACAGATCTTCGAGTTACAGC 1140
1081 CAGTCAATGACTTATGGAATTTAAATGGCTAAACAGTGTGACAGATCTTCGAGTTACAGC 1140
1141 ATCTGTAAGATGAGAGATGATGATGAGAGAGAGTAAAGTACTCGAAGTGCACAA 1200
1141 ATCTGTAAGATGAGAGATGATGATGAGAGAGAGTAAAGTACTCGAAGTGCACAA 1200
1201 GGAGAGAGAGATCCAGAGTTGAACTAGAACCAACCAATGTTTACGATTTTACG 1260
1201 GGAGAGAGAGATCCAGAGTTGAACTAGAACCAACCAATGTTTACGATTTTACG 1260

1261 AGAGTGAAATTTACTCTGCTGTATCTGACAGTGTCTTATAGCCTTTTACTAAGAAAGAGACC 1320
1261 AGAGTGAAATTTACTCTGCTGTATCTGACAGTGTCTTATAGCCTTTTACTAAGAAAGAGACC 1320
1321 AGTGTCTGTTGACAGAGCTCAAAAATTTGATGTTCAAGAGAGAGATCTTCTTTCTGCAAT 1380
1321 AGTGTCTGTTGACAGAGCTCAAAAATTTGATGTTCAAGAGAGAGATCTTCTTTCTGCAAT 1380
1381 CATTAATTCATTCGATCATGGCATCCAGGCCAGAGATGATACAAAAGAGAGATCATCCA 1440
1381 CATTAATTCATTCGATCATGGCATCCAGGCCAGAGATGATACAAAAGAGAGATCATCCA 1440
1441 ATTATGATGGTGTGAAACCCCTTGTGAACAGAGAGTACTTCCACCTACCTTCCCTCGA 1500
1441 ATTATGATGGTGTGAAACCCCTTGTGAACAGAGAGTACTTCCACCTACCTTCCCTCGA 1500
1501 TATGCAAAAATTAATTAAGAGGAAGAAATGGTGAACATTTTTCGAAGATTAATAGATAGA 1560
1501 TATGCAAAAATTAATTAAGAGGAAGAAATGGTGAACATTTTTCGAAGATTAATAGATAGA 1560
1561 ATAAAACTGTCTGTGAGTGTGTAATTTAAACAAATTTACATTTGATCTGCTGATTTTTC 1620
1561 ATAAAACTGTCTGTGAGTGTGTAATTTAAACAAATTTACATTTGATCTGCTGATTTTTC 1620
1621 TGTGAATTTAGTGAACAGTCAACCATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTC 1680
1621 TGTGAATTTAGTGAACAGTCAACCATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTC 1680
1681 CTCGTGATTAACAAAAGGTCTTTGGAACCTCATCTCATGCAAGACATGGTGAAGATGCA 1740
1681 CTCGTGATTAACAAAAGGTCTTTGGAACCTCATCTCATGCAAGACATGGTGAAGATGCA 1740
1741 CTTTCGGTCTTTCTCAGATCTCCAGAGTCTTTCCCGGAGTCTACCTATATAAATTC 1800
1741 CTTTCGGTCTTTCTCAGATCTCCAGAGTCTTTCCCGGAGTCTACCTATATAAATTC 1800
1801 ACCAGGTAAGGACTGTATGACTCTCTTGTACTCACTGTGTGGCCATCTCTGATGC 1860
1801 ACCAGGTAAGGACTGTATGACTCTCTTGTACTCACTGTGTGGCCATCTCTGATGC 1860
1861 TTATTCAGATCCATGACATACAGAGCTCGACAGAGATAGCTTGGTGCATATCTTTC 1920
1861 TTATTCAGATCCATGACATACAGAGCTCGACAGAGATAGCTTGGTGCATATCTTTC 1920
1921 AGGAATTTGCACTCTGAGAGTGTATGACATTTTATTTAATAGGACAGAGAGGT 1980
1921 AGGAATTTGCACTCTGAGAGTGTATGACATTTTATTTAATAGGACAGAGAGGT 1980
1981 TGATGAGCGCTTCACACCATGTGTGAAACAGGAGACCCAGAGGACATTTGGCCTG 2040
1981 TGATGAGCGCTTCACACCATGTGTGTGAAACAGGAGACCCAGAGGACATTTGGCCTG 2040
2041 TTTAGGTACCTGGGTCTTTTACCATTAACCTTGCATTTATGATACAGTACCTTCTAAGTGG 2100
2041 TTTAGGTACCTGGGTCTTTTACCATTAACCTTGCATTTATGATACAGTACCTTCTAAGTGG 2100
2101 CTTTGAATTTGGAATCTTACAGTATGACAGTACTTATACATATTTGATCTCTCTGA 2160
2101 CTTTGAATTTGGAATCTTACAGTATGACAGTACTTATACATATTTGATCTCTCTGA 2160
2161 ATTCTCTTACGCAATGGTGTGATGTCAACATTTGAGTGTGGCTCTCAATGGCAGA 2220
2161 ATTCTCTTACGCAATGGTGTGATGTCAACATTTGAGTGTGGCTCTCAATGGCAGA 2220
2221 GGAAGGATTAATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
2221 GGAAGGATTAATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
2281 GGAAGGATTAATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
2281 GGAAGGATTAATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
2341 TGGAAATTTTAAAAACCATGTTGATAGTATTTGACATGGAGCGGCAAGTACGTTAAACCGAAGTT 2400

Db 2341 TGGATGTTAAACCAATGAGTACATTTGACATGGAGCGCAAGTACGTAAACCGAAGT 2400
Qy 2401 TGAGCTTGATAGTGAACAAGTTCGGTATGAAACACAGGTTTGTCTCAATCAACAGTGTGAT 2460
Db 2401 TGAGCTTGATAGTGAACAAGTTCGGTATGAAACACAGGTTTGTCTCAATCAACAGTGTGAT 2460
Qy 2461 GACCCGCCGCCAGTGCACACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAG 2520
Db 2461 GACCCGCCGCCAGTGCACACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAG 2520
Qy 2521 CCCTCTCTCAGTCTCTGAACTGTATGTGGCAGCTAGTAGCACTTTCAACAGGCAAA 2580
Db 2521 CCCTCTCTCAGTCTCTGAACTGTATGTGGCAGCTAGTAGCACTTTCAACAGGCAAA 2580
Qy 2581 AATGATATTCGAAATATCTTAAACCGGACCACTAGAGTTAATAGAAATTTTAAAGTTGC 2640
Db 2581 AATGATATTCGAAATATCTTAAACCGGACCACTAGAGTTAATAGAAATTTTAAAGTTGC 2640
Qy 2641 CAAACCAACTTTGTGTTATGAAGTTATTTGGCAGGAGGACACAAAGAAATCTTAAAGT 2700
Db 2641 CAAACCAACTTTGTGTTATGAAGTTATTTGGCAGGAGGACACAAAGAAATCTTAAAGT 2700
Qy 2701 TCCTCTGAAATTTGATTTCTCTCATATAATATTTTCTGTTGTGAACTTTGTTGAGA 2760
Db 2701 TCCTCTGAAATTTGATTTCTCTCATATAATATTTTCTGTTGTGAACTTTGTTGAGA 2760
Qy 2761 GAGACTGGGAGTGGCCATAAAGGGGACAGTCTTCTTCAGACCCAACTCTTAGAGGG 2820
Db 2761 GAGACTGGGAGTGGCCATAAAGGGGACAGTCTTCTTCAGACCCAACTCTTAGAGGG 2820
Qy 2821 CACATCACAGGCTCCACATCACGGGAAGTGAATGATTTCTGGGTAACTCACTTAA 2880
Db 2821 CACATCACAGGCTCCACATCACGGGAAGTGAATGATTTCTGGGTAACTCACTTAA 2880
Qy 2881 TAAGGAATCTTTAGTTTTCAGACCCCTTATATGACATGAAATGAAATCTCTGTTTAAAG 2940
Db 2881 TAAGGAATCTTTAGTTTTCAGACCCCTTATATGACATGAAATGAAATCTCTGTTTAAAG 2940
Qy 2941 TGGTTTATATGTTCCATGAAGAACTGGTCTTATGATGATGATGATGATGATGAT 3000
Db 2941 TGGTTTATATGTTCCATGAAGAACTGGTCTTATGATGATGATGATGATGATGAT 3000
Qy 3001 GGTTTTATCAGATTTAATCACAATCATTTTTTATGATGATGATGATGATGATGAT 3060
Db 3001 GGTTTTATCAGATTTAATCACAATCATTTTTTATGATGATGATGATGATGATGAT 3060
Qy 3061 TTATAAGCTTAATAAATTTCTTCACAAAAA 3096
Db 3061 TTATAAGCTTAATAAATTTCTTCACAAAAA 3096

Not Prior Art
Provisional
Appl do not
disclose
"20725"

RESULT 2
US-09-044733-20725
Sequence 20725, Application US/09814353
Publication No. US20030165831A1
GENERAL INFORMATION:
APPLICANT: Lee, John
APPLICANT: Thompson, Pamela
APPLICANT: Lillie, James
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
FILE REFERENCE: MRI-006B
CURRENT APPLICATION NUMBER: US/09/814,353
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/191,031
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: US 60/207,124
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: US 60/211,940
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: US 60/216,820

PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: US 60/220,561
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/257,672
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 22037
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 20725
LENGTH: 2610
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1, 2, 3, 4
OTHER INFORMATION: n = A, T, C or G
US-09-814-353-20725
Query Match 72.08; Score 2228.4; DB 12; Length 2610;
Best Local Similarity 97.98; Pred. No. 0;
Matches 2299; Conservative 0; Mismatches 21; Indels 28; Gaps 3;
Qy 426 CGCATGGGTGACGCTCCGGTCCGGCTGAGAGGGAGGGCGGCGCGCCGCA 485
Db 2610 CGCATGGGTGACGCTCCGGTCCGGCTGAGAGGGAGGGCGGCGCGCCGCA 2551
Qy 486 GCGCGCTGCTTATTTCCGTGCTCCGACAGTCCGTGGCGCGCGGTGACACGCGAGA 545
Db 2550 GCGCGCTGCTTATTTCCGTGCTCCGACAGTCCGTGGCGCGCGGTGACACGCGAGA 2491
Qy 546 AGTAGGCATATGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGAGCTCAG 605
Db 2490 AGTAGGCATATGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGAGCTCAG 2431
Qy 606 TATGCCAGAAAAATGAGAAAGCAATACAACTGGGTGCAATACCCCAAGATTGA 665
Db 2430 TATGCCAGAAAAATGAGAAAGCAATACAACTGGGTGCAATACCCCAAGATTGA 2371
Qy 666 AGAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCTATTTGGTCTTTT 725
Db 2370 AGAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCTATTTGGTCTTTT 2311
Qy 726 TGAAGCCATGCTCTATGAAATGATGGATGCCAAGATGGATGCCATGATTTGAAA 785
Db 2310 TGAAGCCATGCTCTATGAAATGATGGATGCCAAGATGGATGCCATGATTTGAAA 2251
Qy 786 CCAAGTTAATCGAAAGTCTCAATTTTGAACAGCTATCAAGATGGCACTATTAAAT 845
Db 2250 CCAAGTTAATCGAAAGTCTCAATTTTGAACAGCTATCAAGATGGCACTATTAAAT 2191
Qy 846 TAAAGATCTACCTTGGCTGACATGATGGATGATGATGATGATGATGATGATGAT 905
Db 2190 TAAAGATCTACCTTGGCTGACATGATGGATGATGATGATGATGATGATGATGAT 2131
Qy 906 AACGTGGTTAGAGCCATTCACCTGGCAGACAGATGATTTACGTGCTTTACATTCATA 965
Db 2130 AACGTGGTTAGAGCCATTCACCTGGCAGACAGATGATTTACGTGCTTTACATTCATA 2071
Qy 966 TCCAGACTTTATAGAGATCCTGCTATGAAAGGCTTTTGTCTGGAAATCTTGAATCTG 1025
Db 2070 TCCAGACTTTATAGAGATCCTGCTATGAAAGGCTTTTGTCTGGAAATCTTGAATCTG 2011
Qy 1026 TGACATTCAGAGGAAAGTAAATGATGCTGCTTTTGAAGAGAGAGATTTTCAGTC 1085
Db 2010 TGACATTCAGAGGAAAGTAAATGATGCTGCTTTTGAAGAGAGAGATTTTCAGTC 1951
Qy 1086 AATGACTTTATGAAATTTAAATGGCTAAACAGTGGACAGATCTTCGAGTACAGGCTATG 1145
Db 1950 AATGACTTTATGAAATTTAAATGGCTAAACAGTGGACAGATCTTCGAGTACAGGCTATG 1891
Qy 1146 AAAAGATGTGAGGATGACATGCAAGAGAGTAAAGTACTCTGAGTCTGCAAGGAGA 1205
Db 1890 AAAAGATGTGAGGATGACATGCAAGAGAGTAAAGTACTCTGAGTCTGCAAGGAGA 1831

| | | |
|------|-----------------------------------------------------------------|------|
| 1206 | AGAAAGAGATCCAGAAGTTGAACCTAGAAACCAACAAATGTTTAGCAGTATTTCAGCAGAGT | 1265 |
| QY | | |
| 1830 | AGAAAGAGATCCAGAAGTTGAACCTAGAAACCAACAAATGTTTAGCAGTATTTCAGCAGAGT | 1771 |
| Db | | |
| 1266 | GAATTTTACTCGTGTGTTACTGACAGTGCCTTATAGCCCTTACTAAGAAGAGACCAAGTGC | 1325 |
| QY | | |
| 1770 | GAATTTTACTCGTGTGTTACTGACAGTGCCTTATAGCCCTTACTAAGAAGAGACCAAGTGC | 1711 |
| Db | | |
| 1326 | TGTTGCAGAAGCTCAAAAATTGATGGTTCCAGCAGCAGAGATCTTTCTTCTGCCATTCATAA | 1385 |
| QY | | |
| 1710 | TGTTGCAGAAGCTCAAAAATTGATGGTTCCAGCAGCAGAGATCTTTCTTCTGCCATTCATAA | 1651 |
| Db | | |
| 1386 | TTCAATTGCATCATGGCATCCAGGCCCAGAGAATGATCTACAAAGAGAGATCATCCAAATTAT | 1445 |
| QY | | |
| 1650 | TTCAATTGCATCATGGCATCCAGGCCCAGAGAATGATCTACAAAGAGAGATCATCCAAATTAT | 1591 |
| Db | | |
| 1446 | GATGGGTTTTGAACCCCTTGTGAACAGAGGGTACTTTCACACTCCTTCCCTCCGATATGC | 1505 |
| QY | | |
| 1590 | GATGGGTTTTGAACCCCTTGTGAACAGAGGGTACTTTCACACTCCTTCCCTCCGATATGC | 1531 |
| Db | | |
| 1506 | AAAAATAAATAAAGGGAAGAAATGCTGGAACATAATTTTGCAAGATTAATAGATAGAAATAA | 1565 |
| QY | | |
| 1530 | AAAAATAAATAAAGGGAAGAAATGCTGGAACATAATTTTGCAAGATTAATAGATAGAAATAA | 1471 |
| Db | | |
| 1566 | AACGTGCTGTGAGGTTGTGAATTTACAAATTTACATCTGATCTCCTGGAATTTTCTGTGCA | 1625 |
| QY | | |
| 1470 | AACGTGCTGTGAGGTTGTGAATTTACAAATTTACATCTGATCTCCTGGAATTTTCTGTGCA | 1411 |
| Db | | |
| 1626 | ATTTAGTGAACAGTCCACATGTCTTTTCAAGATCTCTGTTACAAACCACTTCTCTGGT | 1685 |
| QY | | |
| 1410 | ATTTAGTGAACAGTCCACATGTCTTTTCAAGATCTCTGTTACAAACCACTTCTCTGGT | 1351 |
| Db | | |
| 1686 | GGATACAAAAAAGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAAAGATGCACTTCG | 1745 |
| QY | | |
| 1350 | GGATACAAAAAAGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAAAGATGCACTTCG | 1291 |
| Db | | |
| 1746 | GTCTTTTGTCCAGATCCCTCCGAGTGCCTTTCCCCCAAGTGTACTATATATAATCAACCAG | 1805 |
| QY | | |
| 1290 | GTCTTTTGTCCAG--TCCTCCGGTGCTTTCCCCCAAGTGTACTATATATAATCAACCAG | 1233 |
| Db | | |
| 1806 | GCTAAGGACGTATCGACTCCTTTGTTACTCACTGTGTTCCGGCCATCTCTGTAGTCTTAT | 1865 |
| QY | | |
| 1232 | GCTAAGGACGTATCGACTCCTTTGTTACTCACTGTGTTCCGGCCATCTCTGTAGTCTTAT | 1173 |
| Db | | |
| 1866 | CAGATCCATCGACATACACGGGCTCGACAGAGATAAGCTTGGTCATTTCTTGAGGAA | 1925 |
| QY | | |
| 1172 | CAGATCCATCGACATACACGGGCTCGACAGAGATAAGCTTGGTCATTTCTTGAGGAA | 1113 |
| Db | | |
| 1926 | TTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTTAAATAGGCAGAGAAGTTGATG | 1985 |
| QY | | |
| 1112 | TTTGCCACCTTGCAGGATG-----AGGCAGAGAAGTTTGTATG | 1076 |
| Db | | |
| 1986 | CAGCGCTTCACACATCGTGTGAAACAGAAACCCCAAGGCAACATTTGGCCTGTTAG | 2045 |
| QY | | |
| 1075 | CAGCGCTTCACACATCGTGTGAAACAGAAACCCCAAGGCAACATTTGGCCTGTTAG | 1016 |
| Db | | |
| 2046 | GTACTGGGTCTTTTACATAACCTTCGCATTTATGATACAGTACCTTCTAAGTGGCTTATG | 2105 |
| QY | | |
| 1015 | GTACTGGGTCTTTTACATAACCTTCGCATTTATGATACAGTACCTTCTAAGTGGCTTATG | 956 |
| Db | | |
| 2106 | AATTGGAACTCTACAGTATGACAGGTAATTTACATATATTGGTATCTCTCGAATTC | 2165 |
| QY | | |
| 955 | AATTGGAACTCTACAGTATGACAGGTAATTTACATATATTGGTATCTCTCGAATTC | 896 |
| Db | | |
| 2166 | TTTAGCATGTTTGATGTCAACATTTAGTCTGGTCCGATGGCTCTCAATGGCAGAGGAAA | 2225 |
| QY | | |
| 895 | TTTAGCATGTTTGATGTCAACATTTAGTCTGGTCCGATGGCTCTCAATGGCAGAGGAAA | 836 |
| Db | | |
| 2226 | GGATATGGACAGCAGCAAGAGGCGGTAGTATRAAAAAACAAGAAAAAAGAAAG | 2285 |
| QY | | |
| 835 | GGATATGGACAGCAGCAAGAGGCGGTAGTATRAAAAAACAAGAAAAAAGAAAG | 776 |
| Db | | |
| 2286 | TTGCGCCATTGAGCGCGAGATCAATATGACCAAGCATATACGAAATGCTGTCTGCTGAA | 2345 |
| QY | | |

| | | | |
|----|------|------------------------------------------------------------------|------|
| Db | 775 | TTCCGCATTGAGCGGAGAGATCACAAATGAGCCAAAGCATATTCAGAACATGTGTCTGTGAA | 7116 |
| QY | 2346 | TGTTTAAAAACCATGGTAGCATTTGCATCGGACGGCAAAAGTACGTAAACCGGAAGTTTGAGC | 2405 |
| Db | 715 | TGTTTAAAAACCATGGTAGCATTTGCATCGGACGGCAAAAGTACGTAAACCGGAAGTTTGAGC | 656 |
| QY | 2406 | TTGATAGTGAACAAGTTCGGTATGAAACACAGAGTTTGCTCCATTCAAACAGTGTGATGACCC | 2465 |
| Db | 655 | TTGATAGTGAACAAGTTCGGTATGAAACACAGAGTTTGCTCCATTCAAACAGTGTGATGACCC | 596 |
| QY | 2466 | CGCCGCACGTGCACACTCTTACAGTTCAAAGAAATGTCGACCTCAATAAATATGACCCCTC | 2525 |
| Db | 595 | CGCCGCACGTGCACACTCTTACAGTTCAAAGAAATGTCGACCTCAATAAATATGACCCCTC | 536 |
| QY | 2526 | CTCCTCAGTCTCCTGGAACCTGTATGTGGCAGCTAGTAAGCACCTTCCACACAGGCCAAAATGA | 2585 |
| Db | 535 | CTCCTCAGTCTCCTGGAACCTGTATGTGGCAGCTAGTAAGCACCTTCCACACAGGCCAAAATGA | 476 |
| QY | 2586 | TATTGGAAAAATTTCCCTAAACCCGGACCATGAGGTTAATAGATTTTAAAGGTCGCCAAC | 2645 |
| Db | 475 | TATTGGAAAAATTTCCCTAAACCCGGACCATGAGGTTAATAGATTTTAAAGGTCGCCAAC | 416 |
| QY | 2646 | CCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTTAAAGTTCCTC | 2705 |
| Db | 415 | CCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTTAAAGTTCCTC | 356 |
| QY | 2706 | CTGATTTTGAATTTCTCTGCTCATAAATATTTTCTCTGTGAAACTGTGTTTGAGAGAGAC | 2765 |
| Db | 355 | CTGAATTTGAATTTCTCTGCTCATAAATTTTCTCTGTGAAACTGTGTTTGAGAGAGACTG | 299 |
| QY | 2766 | TGGGAGG | 2773 |
| Db | 298 | GGAGGTGG | 291 |

RESULT 3

```

US-09-814-353-15927/c
; Sequence 15927, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15927
; LENGTH: 745
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 571
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-15927

```


Query Match 18.1%; Score 560.8; DB 12; Length 745;
Best Local Similarity 94.8%; Pred. No. 2.2e-141;
Matches 621; Conservative 0; Mismatches 8; Indels 26; Gaps 3;
QY 1399 GGCATCCAGGCCAGAGATGATACACAAAGGAGATCATCCAAATATGATGGTTTGA 1458
DB 739 GGCATCCAGGCCAGAGATGATACACAAAGGAGATCATCCAAATATGATGGTTTGA 680
QY 1459 CCCCTTGTGAACGAGGCTACTTCCACCTACCTCCCTCGATATGCAAAATATATAA 1518
DB 679 CCCCTTGTGAACGAGGCTACTTCCACCTACCTCCCTCGATATGCAAAATATATAA 620
QY 1519 AGGGAAGAAATGGTGAATCTTTTGAAGATTAATAGATAGATAAAACTCTCTGTGAG 1578
DB 619 AGGGAAGAAATGGTGAATCTTTTGAAGATTAATAGATAGATAAAANCTCTGTGAG 560
QY 1579 GTTGTGAATTAACAAATTTACATTTGATCTCTGGA-TTTTCTGTGAAATTAAGTGA 1637
DB 559 GTTGTGAATTAACAAATTTACATTTGATCTCTGGAATTTTCTGTGAAATTAAGTGA 500
QY 1638 GTCACCATGTGTTCTTTCAAGATCTCTGTTCACAAACACCTTCCCTGGTGATACAA 1697
DB 499 GTCACCATGTGTTCTTTCAAGATCTCTGTTCACAAACACCTTCCCTGGTGATACAA 440
QY 1698 GGTCTTTGGAACTCATCTGCAAGACATGTTGAAGATGCACTTCGGTCTTTTGTGAG 1757
DB 439 GGTCTTTGGAACTCATCTGCAAGACATGTTGAAGATGCACTTCGGTCTTTTGTGAG 380
QY 1758 ATCTCTCCAGAGCTTTCCCAAGTGTCTATATATAATACACAGGCTAAGGACTGT 1817
DB 379 TCCTCCG-GTGTCTTTCCCAAGTGTCTATATATAATACACAGGCTAAGGACTGT 322
QY 1818 ATGCACTCTTTGTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1877
DB 321 ATGCACTCTTTGTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 262
QY 1878 CATACAGGGCTCGACAGAGATAGCTTTGGTCTATATCTTGAAGAAATTCGACCTTG 1937
DB 261 CATACAGGGCTCGACAGAGATAGCTTTGGTCTATATCTTGAAGAAATTCGACCTTG 202
QY 1938 CAGGATGATTTATGACATTTTATTTAATAGCAGAGAGGTTGATGAGGCTTCACA 1997
DB 201 CAGGATG-----AGGAGAGAGGTTGATGAGGCTTCACA 165
QY 1998 CCATGCTGTTGAACAGGAACCCCAAGGCAACATTTGGCTGTTTGTAGTACCTG 2052
DB 164 CCATGCTGTTGAACAGGAACCCCAAGGCAACATTTGGCTGTTTGTAGTACCTG 110

RESULT 4

US-09-814-353-16155
; Sequence 16155, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661

; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16155
; LENGTH: 610
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-16155
Query Match 16.5%; Score 509.4; DB 12; Length 610;
Best Local Similarity 99.8%; Pred. No. 1.8e-127;
Matches 510; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1184 GTACTCGAAGTCGACAGAGAGAGATCCAGAGATTGAACTAGAACACCAACAA 1243
DB 100 GTACTCGAAGTCGACAGAGAGAGATCCAGAGATTGAACTAGAACACCAACAA 159
QY 1244 GTTACAGATATTCAGACAGAGTGAATTTACTTCGTGTGTACTGACAGTGTATAGCT 1303
DB 160 GTTACAGATATTCAGACAGAGTGAATTTACTTCGTGTGTACTGACAGTGTATAGCT 219
QY 1304 TTACTAAGAAAGACACAGTGTCTGTGCAAGAGCTCAAAAATTTGATGGTTCAAGCAG 1363
DB 220 TTACTAAGAAAGACACAGTGTCTGTGCAAGAGCTCAAAAATTTGATGGTTCAAGCAG 279
QY 1364 ATCTCTTCTGCAATTCATTAATTCATTCATTCATTCATTCATTCATTCATTCATTC 1423
DB 280 ATCTCTTCTGCAATTCATTAATTCATTCATTCATTCATTCATTCATTCATTCATTC 339
QY 1424 CAAAAGAGATCATCAATTTATGATGGTGTGAAACCCCTTCTGAAACAGAGGCTACTTC 1483
DB 340 CAAAAGAGATCATCAATTTATGATGGTGTGAAACCCCTTCTGAAACAGAGGCTACTTC 399
QY 1484 CACTACTCTTCCCTCGATATGCAAAATTAATTAAGGGAAGAAATGTTGAACTATTTTG 1543
DB 400 CACTACTCTTCCCTCGATATGCAAAATTAATTAAGGGAAGAAATGTTGAACTATTTTG 459
QY 1544 CAAGATTAATAGATAGATAAATAAATCTGTGTGAGTGTGTAATTAACAAATTTTACATT 1603
DB 460 CAAGATTAATAGATAGATAAATAAATCTGTGTGAGTGTGTAATTAACAAATTTTACATT 519
QY 1604 GTATCTCGAATTTTCTGTGAAATTTAGTGAACAGTCACCP TGTGTCTTTTCAAGATCTC 1663
DB 520 GTATCTCGAATTTTCTGTGAAATTTAGTGAACAGTCACCP TGTGTCTTTTCAAGATCTC 579
QY 1664 TGTTCACAAACCACTTTCTCTGTTGATAACAA 1694
DB 580 TGTTCACAAACCACTTTCTCTGTTGATAACAA 610

RESULT 5

US-09-814-353-3218/c
; Sequence 3218, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820

```

; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3218
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 561..590
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-3218

Query Match          16.0%; Score 496.2; DB 12; Length 602;
Best Local Similarity 93.9%; Pred. No. 6.8e-124;
Matches 568; Conservative 0; Mismatches 10; Indels 27; Gaps 4;

QY 1450 GGTTTGAACCCCTTGTGAACAGAGGCTACTTCCACCTACCTTCCCTCGATATGC-AAA 1508
DB 602 GGTTTGAACCCCTTGTGAACAGAGGCTACTTCCACCTACCTTCCCTCGATATGC-AAA 543
QY 1509 AATAATTAAAGGGAAGAAATGGTGAACCTATTTTGCAGGATTAATAGATAGATAAAAC 1568
DB 542 AATAATTAAAGGGAAGAAATGGTGAACCTATTTTGCAGGATTAATAGATAGATAAAAC 483
QY 1569 TGTCTGTGAGGTTGTGAATTTAAACAAATTTACATGTATCTCTGA-TTTTTCTGTGAAT 1627
DB 482 TGTCTGTGAGGTTGTGAATTTAAACAAATTTACATGTATCTCTGAATTTTTTCTGTGAAT 423
QY 1628 TTAGTGAACAGTCCACCTGTCTTCTTCAAGATCTCTGTTACAAACCCACTTTCCTGTGG 1687
DB 422 TTAGTGAACAGTCCACCTGTCTTCTTCAAGATCTCTGTTACAAACCCACTTTCCTGTGG 363
QY 1688 ATAACAAAAGGCTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCATTCGGT 1747
DB 362 ATAACAAAAGGCTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCATTCGGT 125
QY 1748 TGCCACCTTGAGGATGATTTATGACATTTTATTTTAAATAGGACAGAGAGTTGATGCA 1987
DB 124 TGCCACCTTGAGGATG-----AGGCAGAGAGGTTGATGCA 88
QY 1988 GCGCTTCACCATGCTGTTGAAACAGGAAACCCCAAGGCAACATTTGGCCTGTTAGGT 2047
DB 87 GCGCTTCACCATGCTGTTGAAACAGGAAACCCCAAGGCAACATTTGGCCTGTTAGGT 28
QY 2048 ACCTG 2052
DB 27 ACCTG 23

RESULT 6
US-09-814-353-9543/c
; Sequence 9543, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela

```

Db 87 GCGCTTACACCATGCTGTGTAAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAGGT 28
 QY 2048 ACTGT 2052
 Db 27 ACTGT 23

RESULT 7

US-09-918-995-20914
 ; Sequence 20914, Application US/09918995
 ; Publication No. US20030073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; FROM VARIOUS CDNA LIBRARIES
 ; FILE REFERENCE: 20411-756
 ; CURRENT APPLICATION NUMBER: US/09/918,995
 ; PRIOR FILING DATE: 2001-07-30
 ; PRIOR APPLICATION NUMBER: US/09/235,076
 ; PRIOR FILING DATE: 1999-01-20
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 20914
 ; LENGTH: 482
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)...(482)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-918-995-20914

Query Match 14.6%; Score 451.2; DB 11; Length 482;
 Best Local Similarity 99.3%; Pred. No. 9.9e-112;
 Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 491 GTCGTATTTCGGTCCGACAGTGCCTGGCGCGGGTGACCCCGGAGAGTAG 550
 Db 26 CGAGCTATTTCGGTCCGACAGTGCCTGGCGCGGGTGACCCCGGAGAGTAG 85
 QY 551 GCATAATGGTTATGAAGCTTCTGTAGATGATGACGATTCAGATGGGAGCTCAGTAGC 610
 Db 86 GCATAATGGTTATGAAGCTTCTGTAGATGATGACGATTCAGATGGGAGCTCAGTAGC 145
 QY 611 CAGAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATACCCAAAGATTTGAAGAAG 670
 Db 146 CAGAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATACCCAAAGATTTGAAGAAG 205
 QY 671 CTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAGCTATTTGGTCTTTTGAAG 730
 Db 206 CTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAGCTATTTGGTCTTTTGAAG 265
 QY 731 CCATGCTGCTATTGAAATGATGGATCCCAAGATGGATCTGGCATGATTGGAACCAAG 790
 Db 266 CCATGCTGCTATTGAAATGATGGATCCCAAGATGGATCTGGCATGATTGGAACCAAG 325
 QY 791 TTAATCGAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAATAATTAAG 850
 Db 326 TTAATCGAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAATAATTAAG 385
 QY 851 ATCTCACCTTGGCTGAACTGATAGGATATGATACATGTTTTTCTGTTTGAACGT 910
 Db 386 ATCTCACCTTGGCTGAACTGATAGGATATGATACATGTTTTTCTGTTTGAACGT 445
 QY 911 GGTTAGAGGCCATTCACTGGCACAGCAGTATTTA 946
 Db 446 GGTTAGAGGCCATTCACTGGCACAGCAGTATTTA 481

RESULT 8

US-09-814-353-3455
 ; Sequence 3455, Application US/09814353
 ; Publication No. US20030165831A1

; GENERAL INFORMATION:
 ; APPLICANT: Lee, John
 ; APPLICANT: Thompson, Pamela
 ; APPLICANT: Lillie, James
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
 ; IDENTIFICATION, ASSESSMENT, PREVENTION, AND
 ; THERAPY OF OVARIAN CANCER
 ; FILE REFERENCE: MRI-006B
 ; CURRENT APPLICATION NUMBER: US/09/814,353
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: US 60/191,031
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: US 60/207,124
 ; PRIOR FILING DATE: 2000-05-25
 ; PRIOR APPLICATION NUMBER: US 60/211,940
 ; PRIOR FILING DATE: 2000-06-15
 ; PRIOR APPLICATION NUMBER: US 60/216,820
 ; PRIOR FILING DATE: 2000-07-07
 ; PRIOR APPLICATION NUMBER: US 60/220,661
 ; PRIOR FILING DATE: 2000-07-25
 ; PRIOR APPLICATION NUMBER: US 60/257,672
 ; PRIOR FILING DATE: 2000-12-21
 ; NUMBER OF SEQ ID NOS: 22037
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3455
 ; LENGTH: 434
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-814-353-3455

Query Match 14.0%; Score 434; DB 12; Length 434;
 Best Local Similarity 100.0%; Pred. No. 4.3e-107;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1186 ACTCGAAGTCGACAGGAGAGAGATCCAGAGTTGAACACTAGAACACCAATGT 1245
 Db 1 ACTCGAAGTCGACAGGAGAGAGATCCAGAGTTGAACACTAGAACACCAATGT 60
 QY 1246 TTAGCAGTATTTCAGCAGAGTGAAATTTACTCGTGTGTACTGCACAGTGTCTATAGCCTTT 1305
 Db 61 TTAGCAGTATTTCAGCAGAGTGAAATTTACTCGTGTGTACTGCACAGTGTCTATAGCCTTT 120
 QY 1306 ACTAAGAAAGAGACAGTGTCTGTGAGAGCTCAAAAATGATGTTCAAGCAGAGAT 1365
 Db 121 ACTAAGAAAGAGACAGTGTCTGTGAGAGCTCAAAAATGATGTTCAAGCAGAGAT 180
 QY 1366 CTCTCTTCTGCCATTTCATAATTCATTGTCATCCAGCCCGCAGATGATACTACA 1425
 Db 181 CTCTCTTCTGCCATTTCATAATTCATTGTCATCCAGCCCGCAGATGATACTACA 240
 QY 1426 AAAGGAGATCATCAATTAATGATGGGTTTGAACCCCTTGTGAACAGAGGCTACTTCCA 1485
 Db 241 AAAGGAGATCATCAATTAATGATGGGTTTGAACCCCTTGTGAACAGAGGCTACTTCCA 300
 QY 1486 CCTACCTTCCCTCGATATGCAAAAATAATTAAGGAGAGAAATGTTGAACCTATTTTTGA 1545
 Db 301 CCTACCTTCCCTCGATATGCAAAAATAATTAAGGAGAGAAATGTTGAACCTATTTTTGA 360
 QY 1546 AGATTAAATAGATAGATAAAAACCTGCTGTGAGGTGTGAATTTAACAAATTTACATGT 1605
 Db 361 AGATTAAATAGATAGATAAAAACCTGCTGTGAGGTGTGAATTTAACAAATTTACATGT 420
 QY 1606 ATCCTGGATTTTTT 1619
 Db 421 ATCCTGGATTTTTT 434

RESULT 9

US-09-814-353-9771
 ; Sequence 9771, Application US/09814353
 ; Publication No. US20030165831A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lee, John

```

; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9771
; LENGTH: 434
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-814-353-9771

```

| Query Match | 14.0% | Score 434 | DB 12 | Length 434 |
|-----------------------|----------------|------------------------------------------------------------------|----------|------------|
| Best Local Similarity | 100.0% | Prod. No. 4.3e-107 | | |
| Matches 434 | Conservative 0 | Mismatches 0 | Indels 0 | Gaps 0 |
| QY | 1186 | ACTCGAAGTCGACAAAGGAGAGAAAGAGATCCAGAAAGTTGAACTAGAAACACCAACAATGT | 1245 | |
| DB | 1 | ACTCGAAGTCGACAAAGGAGAGAAAGAGATCCAGAAAGTTGAACTAGAAACACCAACAATGT | 60 | |
| QY | 1246 | TTAGCAGATATTCAGCAGAGTGAATTTACTCGTGTGTACTGACAGTGCCTTATAGCCTTT | 1305 | |
| DB | 61 | TTAGCAGATATTCAGCAGAGTGAATTTACTCGTGTGTACTGACAGTGCCTTATAGCCTTT | 120 | |
| QY | 1306 | ACTAAGAAAGAGACCCAGTGTGTTCGAGAAGCTCAAAAATTTGATGTTCAAGCAGACAGAT | 1365 | |
| DB | 121 | ACTAAGAAAGAGACCCAGTGTGTTCGAGAAGCTCAAAAATTTGATGTTCAAGCAGACAGAT | 180 | |
| QY | 1366 | CTTCTTTCTGCCATTCAATAATTCATGTCATCATGGCATCCAGGCCCAGAAATGATCTACA | 1425 | |
| DB | 181 | CTTCTTTCTGCCATTCAATAATTCATGTCATCATGGCATCCAGGCCCAGAAATGATCTACA | 240 | |
| QY | 1426 | AAAGGAGATCATCCAATTATGATGGGTTTTGAACCCCTCTGTAACACAGAGGCTACTTCCA | 1485 | |
| DB | 241 | AAAGGAGATCATCCAATTATGATGGGTTTTGAACCCCTCTGTAACACAGAGGCTACTTCCA | 300 | |
| QY | 1486 | CCTACCTTCCTCGATATGCAAAAATAATTTAAAGGGAGAAATGGTGAACTATTTGCA | 1545 | |
| DB | 301 | CCTACCTTCCTCGATATGCAAAAATAATTTAAAGGGAGAAATGGTGAACTATTTGCA | 360 | |
| QY | 1546 | AGATTAAATAGATAGATAAAAACCTGCTGAGAGGTTGTTGAAATTAACAAATTTACATTTGT | 1605 | |
| DB | 361 | AGATTAAATAGATAGATAAAAACCTGCTGAGAGGTTGTTGAAATTTAAACAAATTTACATTTGT | 420 | |
| QY | 1606 | ATCCTGGATTTTTT | 1619 | |
| DB | 421 | ATCCTGGATTTTTT | 434 | |

RESULT 10
US-09-983-965-210
: Sequence 210, Application US/09983965
: Patent No. US20020137160A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Ningbing
: APPLICANT: Byatt, John C.
:

```

; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 210
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 25-BOVMS1-018-Q1-E1-G1
US-09-983-965-210

```

| Query Match | 11.3% | Score 350.8; | DB 10; | Length 398; |
|-----------------------|-----------------|---------------------------------------------------------------|-----------|-------------|
| Best Local Similarity | 94.3%; | Pred. No. 1.7e-84; | | |
| Matches 364; | Conservative 0; | Mismatches 22; | Indels 0; | Gaps 0; |
| QY | 2298 | GGCGAGAGATCAAAATGAGCCCAAGCATATCAGAACATGTGCTCGAAATGTTTAAACCA | 2357 | |
| Db | 13 | GGCGAGAGATCAAAATGAGCCCAAGCATATCAGAACATGTGCTCGAAATGTTTAAACCA | 72 | |
| QY | 2358 | TGTTAGCATTTGCATGAGCGGCAAAAGTAGTAAACCGAAGTTTGAGCTTTGATAGTGAAC | 2417 | |
| Db | 73 | TGGGAGCTTTGCATGAGTGCGCAAAAGTAGTAAACCGAAGTTTGAGCTTTGATAGTGAAC | 132 | |
| QY | 2418 | AAGTTGCGTATGAACACAGGTTTGCTCCATTTCAACAGTGTGATGACCCCGCCGAGTGC | 2477 | |
| Db | 133 | AAGTTGCGTATGAGCAGCATTTGCTCCATTTCAACAGTGTGATGACCCCGAGTGC | 192 | |
| QY | 2478 | ACTACTTACAGTTTCAAGGAAATGCTGACCTCAATAAATAGCCCTCCCTCAGTCTC | 2537 | |
| Db | 193 | ACTATCTGCAGTTTCAAGGAAATGCTGACCTCAATAAATAGCCCTCCCTCAGTCTC | 252 | |
| QY | 2538 | CTGAACCTGATGTGGCAGCTAGTAGACACTTTTCAACAGGCAAAAATGATATTGGAAAAA | 2597 | |
| Db | 253 | CAGAACTGATGTGGCAGCTAGTAGACACTTTTCAACAGGCAAAAATGATACTGGAAAAA | 312 | |
| QY | 2598 | TTCCCTAACCCGACCATGAGTTTAATAGATTTTAAAGTTTGCACACCCCAACTTTGTGG | 2657 | |
| Db | 313 | TCCCAACCCAGACCATGAGGTCAATAGAAATTTTAAAGGTTTGCACACCCCAACTTTGTGG | 372 | |
| QY | 2658 | TTATGAAGTTATTGGCAGGAGGACAC | 2683 | |
| Db | 373 | TTATGAAGTTATTGGCAGGAGGACAC | 398 | |

RESULT 11
US-09-783-590-5693
Sequence 5693, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:
APPLICANT: Dillon, Patrick J.
APPLICANT: Haseltine, William A.
APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: PO-16.2C1
CURRENT APPLICATION NUMBER: US/09/783,590
CURRENT FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 08/420,856
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5693

;; TITLE OF INVENTION: Polymorphisms in the Human Genome
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 323193
;; LENGTH: 489
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-323193

Query Match 5.9%; Score 182; DB 13; Length 489;
Best Local Similarity 100.0%; Pred. No. 1.6e-38;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1431 AGATCATCAATTATGATGGGTTTGAACCCCTTGTGAACGAGGCTACTTCCACCTAC 1490
DB |||||||
QY 401 AGATCATCAATTATGATGGGTTTGAACCCCTTGTGAACGAGGCTACTTCCACCTAC 342
DB |||||||
QY 1491 CTTCCTCGATATGCAAAAATAATTAAGGGAAGAAATGGTGAACCTATTTTGCAGATT 1550
DB |||||||
QY 341 CTTCCTCGATATGCAAAAATAATTAAGGGAAGAAATGGTGAACCTATTTTGCAGATT 282
DB |||||||
QY 1551 AATGATGAGATAAAACCTGTCTGTGAGGTTGTGAATTAACAAATTTACATTTGATCCT 1610
DB |||||||
QY 281 AATGATGAGATAAAACCTGTCTGTGAGGTTGTGAATTAACAAATTTACATTTGATCCT 222
DB |||||||
QY 1611 GG 1612
DB ||

RESULT 14

US-10-001-857-41/c
;; Sequence 41, Application US/10001857
;; Publication No. US20020183500A1
;; GENERAL INFORMATION:
;; APPLICANT: Macina, Roberto
;; APPLICANT: Recipon, Herve
;; APPLICANT: Chen, Sei-Yu
;; APPLICANT: Sun, Yongming
;; APPLICANT: Liu, Chenghua
;; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prob
;; FILE REFERENCE: DEX-0273
;; CURRENT APPLICATION NUMBER: US/10/001,857
;; CURRENT FILING DATE: 2001-11-20
;; PRIOR APPLICATION NUMBER: 60/252,054
;; PRIOR FILING DATE: 2000-11-20
;; NUMBER OF SEQ ID NOS: 208
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 41
;; LENGTH: 151
;; TYPE: DNA
;; ORGANISM: Homo sapien
US-10-001-857-41

Query Match 4.5%; Score 140.4; DB 13; Length 151;
Best Local Similarity 99.3%; Pred. No. 1.3e-27;

Matches 141; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1911 CATATTTCTTGAGGAATTTGCCACCTTGCCAGGATGAGTTTATGACATTTTATTTAATAGG 1970
DB |||||||
QY 151 CATATTTCTTGAGGAATTTGCCACCTTGCCAGGATGAGTTTATGACATTTTATTTAATAGG 92
DB |||||||
QY 1971 CAGAGAAGGTTTGATCGAGCGTTTACACCATCTCTGTGAAAAAGGAAACCCCAAGGCAAC 2030
DB |||||||
QY 91 CAGAGAAGGTTTGATCGCGGCTTACACCATCTCTGTGAAACAGGAACCCCAAGGCAAC 32
DB |||||||
QY 2031 ATTTGGCCTGTTTAGGTACCTG 2052
DB |||||||
QY 31 ATTTGGCCTGTTTAGGTACCTG 10
DB |||||||

RESULT 15

US-10-001-857-111/c
;; Sequence 111, Application US/10001857
;; Publication No. US20020183500A1
;; GENERAL INFORMATION:
;; APPLICANT: Macina, Roberto
;; APPLICANT: Recipon, Herve
;; APPLICANT: Chen, Sei-Yu
;; APPLICANT: Sun, Yongming
;; APPLICANT: Liu, Chenghua
;; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P
;; FILE REFERENCE: DEX-0273
;; CURRENT APPLICATION NUMBER: US/10/001,857
;; CURRENT FILING DATE: 2001-11-20
;; PRIOR APPLICATION NUMBER: 60/252,054
;; PRIOR FILING DATE: 2000-11-20
;; NUMBER OF SEQ ID NOS: 208
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 111
;; LENGTH: 815
;; TYPE: DNA
;; ORGANISM: Homo sapien
;; NAME/KEY: misc feature
;; LOCATION: (206)-(305)
;; OTHER INFORMATION: a, c, g or t
US-10-001-857-111

Query Match 2.8%; Score 87.8; DB 13; Length 815;
Best Local Similarity 83.0%; Pred. No. 9.4e-13;
Matches 122; Conservative 0; Mismatches 2; Indels 23; Gaps 1;

QY 1910 TCATATTTCTTGAGGAATTTGCCACCTTGCCAGGATGAGTTTATGACATTTTATTTAATAG 1969
DB |||||||
QY 205 TCATATTTCTTGAGGAATTTGCCACCTTGCCAGGATGAGTTTATGACATTTTATTTAATAG 169
DB |||||||
QY 1970 GCAGAGAAGGTTTGATCGAGCGCTTACACCATCTCTGTGAAACAGGAACCCCAAGGCAAC 2029
DB |||||||
QY 168 GCAGAGAAGGTTTGATCGAGCGCTTACACCATCTCTGTGAAACAGGAACCCCAAGGCAAC 109
DB |||||||
QY 2030 CATTGGCCTGTTTAGGTACCTGGGC 2056
DB |||||||
QY 108 CATTGGCCTGTTTAGGTACCTGGGC 82
DB |||||||

Search completed: November 25, 2003, 03:08:51
Job time : 938 secs

| Result | No. | Query | | | DB | ID | Description |
|--------|-----|--------|-------|--------|----|----------|-----------------|
| | | Score | Match | Length | | | |
| 1 | | 3096 | 100.0 | 3096 | 24 | AB075303 | Human lung |
| C | 2 | 2577.8 | 83.3 | 2668 | 23 | ABV23940 | Human prostate |
| C | 3 | 2577.8 | 83.3 | 2668 | 23 | ABV29823 | Human prostate |
| | 4 | 2417.2 | 78.1 | 2488 | 24 | AAL49929 | Human mole |
| | 5 | 717.4 | 23.2 | 719 | 24 | AB951309 | cDNA encoded |
| | 6 | 350.8 | 11.3 | 398 | 25 | ABX50281 | Bovine EST |
| | 7 | 343.8 | 11.1 | 799 | 24 | ABQ51658 | Oligonucleotide |
| C | 8 | 343.8 | 11.1 | 799 | 24 | ABQ51659 | Oligonucleotide |

XX
PT
New lung specific nucleic acid useful in gene therapy or as vaccines

| | | | |
|------|----|-----------------------------------------------------------------|------|
| 1741 | DB | CTTCGCTCTTTGTGAGATCCCTCCGAGTCTTCTCCCCCAAGTGCCTACCTATATAATAATC | 1800 |
| 1801 | QY | ACGAGCTAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTGGCCATCTCTGTAGTC | 1860 |
| 1801 | DB | ACGAGCTAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTGGCCATCTCTGTAGTC | 1860 |
| 1861 | QY | TTATTCCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATCTTTG | 1920 |
| 1861 | DB | TTATTCCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATCTTTG | 1920 |
| 1921 | QY | AGGAATTCGCCACTTGCAGGATGAGTTTATGACATTTTATTTTAAATAGGCAGAGAAGT | 1980 |
| 1921 | DB | AGGAATTCGCCACTTGCAGGATGAGTTTATGACATTTTATTTTAAATAGGCAGAGAAGT | 1980 |
| 1981 | QY | TGATGCAGCGCTTCACACCATGCTGTGAAACAGGAACCCCAAGGCAACATTTGGCGCTG | 2040 |
| 1981 | DB | TGATGCAGCGCTTCACACCATGCTGTGTGAAACAGGAACCCCAAGGCAACATTTGGCGCTG | 2040 |
| 2041 | QY | TTTAGGTAACCTGGGTCTTTACATAACCTTGCATTTATGATACAGTACCTTCTAAGTGG | 2100 |
| 2041 | DB | TTTAGGTACCTGGGTCTTTACCATAACCTTGCATTTATGATACAGTACCTTCTAAGTGG | 2100 |
| 2101 | QY | CTTTGAATTTGGAACTCTACAGTATGCACGAGTACTATTACATATATTGTTATCTCTCTGA | 2160 |
| 2101 | DB | CTTTGAATTTGGAACTCTACAGTATGCACGAGTACTATTACATATATTGTTATCTCTCTGA | 2160 |
| 2161 | QY | ATTCCCTTTACGCATGGTTGATGTCAACATGTAGTGTGCGGATGGCTCTCAAAATGGCAGA | 2220 |
| 2161 | DB | ATTCCCTTTACGCATGGTTGATGTCAACATGTAGTGTGCGGATGGCTCTCAAAATGGCAGA | 2220 |
| 2221 | QY | GGAAAGGATAATCGRAGACGACAGAAAGGCCGTAGTAGTAAAAAAACAAAGAAAAAAA | 2280 |
| 2221 | DB | GGAAAGGATAATCGRAGACGACAGAAAGGCCGTAGTAGTAAAAAAACAAAGAAAAAAA | 2280 |
| 2281 | QY | GAAAGTTTCGCCCATTTGAGCCGAGAGATCACAATGAGCCCAAGCATATCAAGACATGTGTGC | 2340 |
| 2281 | DB | GAAAGTTTCGCCCATTTGAGCCGAGAGATCACAATGAGCCCAAGCATATCAAGACATGTGTGC | 2340 |
| 2341 | QY | TGGAATGTTTAAAAACCATGTGTAGCATTTGACATGGACGGCAAGTAGCTAAACCGAAGTT | 2400 |
| 2341 | DB | TGGAATGTTTAAAAACCATGTGTAGCATTTGACATGGACGGCAAGTAGCTAAACCGAAGTT | 2400 |
| 2401 | QY | TGAGCTGTAGTAGTGAAACAGTTCGGTAGTAACACAGGTTTGCTTCCATTCAAACAGTGTGAT | 2460 |
| 2401 | DB | TGAGCTGTAGTAGTGAAACAGTTCGGTAGTAACACAGGTTTGCTTCCATTCAAACAGTGTGAT | 2460 |
| 2461 | QY | GACCCGCCCGCAGTGCACTACTTACAGTTCAAGGAAATGCTGTGACCTCAATTAATATAG | 2520 |
| 2461 | DB | GACCCGCCCGCAGTGCACTACTTACAGTTCAAGGAAATGCTGTGACCTCAATTAATATAG | 2520 |
| 2521 | QY | CCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAGAACAATTTTCAACAGGCAAA | 2580 |
| 2521 | DB | CCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAGAACAATTTTCAACAGGCAAA | 2580 |
| 2581 | QY | AATGATATTGGAAATATTCTAACCCGGACCATGAGTTTAAATAGAAATTTTAAAGGTTGC | 2640 |
| 2581 | DB | AATGATATTGGAAATATTCTAACCCGGACCATGAGTTTAAATAGAAATTTTAAAGGTTGC | 2640 |
| 2641 | QY | CAAAACCAAATTTGTGTGTATGAAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGT | 2700 |
| 2641 | DB | CAAAACCAAATTTGTGTGTATGAAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGT | 2700 |
| 2701 | QY | TCCTCCTGAAATTTGATTTCTCTGCTCATAAATATTTTCTGTGTGAAACTCTGTTGAGA | 2760 |
| 2701 | DB | TCCTCCTGAAATTTGATTTCTCTGCTCATAAATATTTTCTGTGTGAAACTCTGTTGAGA | 2760 |
| 2761 | QY | GAGACTGGGAGGTGGCCATAAAGGGCAGAGTCTTCTTCAGACCCCACTCTTAGAGGG | 2820 |
| 2761 | DB | GAGACTGGGAGGTGGCCATAAAGGGCAGAGTCTTCTTCAGACCCCACTCTTAGAGGG | 2820 |
| 2821 | QY | CACATCACCGGCTCCACATCACGGGAAGTGAGATGATTTCTTTGGGTAAACAACATTA | 2880 |

RESULT 2

RESULT 2
ABV23940/C

ID ABV23940 standard; cDNA; 2668 BP.

AA
AC

16-SEP-2002 (first entry)

Human prostate expression marker cDNA 23931.

[illegible]

KW pharmacogenomic marker; gene; ss.

OS ~~Homo sapiens~~.

AX
PN
WO200160860-2

XX PD 23-DEC-2007

1

XX

PR 16-MAR-2000; 2000US-189862P.

PR 25-MAY-2000; 2000US-20/454P.
PR 09-JUN-2000: 2000US-211314P

PR 18-JUL-2000; 2000US-219007P.

XX

XX
XXXXXXXXXXXXXXXXXXXXX / XXXXXX
ИЗДАНИЕ К, БИНАРНОЕ МО, МОИ

DR WPI; 2001-662795/76.
yy -

Novel isolated nucleic acid molecules

PT for detecting presen

PS Claim 1; Page 4423 11750pp; English.

CC XX The invention relates to an isolated

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:

(b) monitoring the progression of prostate cancer in a patient;

1C) assessing the efficacy of a test compound in murine prostate cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient:

(e) selecting a composition for inhibiting prostate cancer in a patient

CC (g) determining whether prostate cancer has metastasized in a patient;

CC patient;

did not designate
us

not Applicable

E MEDICINE INC.
ahan JE;

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 2668 BP; 742 A; 617 C; 483 G; 820 T; 6 other;

SG Query Match 83.3%; Score 2577.8; DB 23; Length 2668;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 2638; Conservative 0; Mismatches 7; Indels 26; Gaps 3;

| | | | |
|----|------|----------------------------------------------------------------|------|
| Qy | 426 | CGCATGCTGCACGCTGCGGCTCGGGCTGGGCTGAGAGGGAGGGGGCGGGCGGCGGA | 485 |
| Db | 2668 | CGCATGCTGCACGCTGCGGCTCGGGCTGGGCTGAGAGGGAGGGGGCGGGCGGCGGA | 2609 |
| Qy | 486 | GGCGGCTGCTTATTTCCGCTGCTCGGACAGTGCCTGGCGGCGGGTGACACAGGAGA | 545 |
| Db | 2608 | GGCGGCTGCTTATTTCCGCTGCTCGGACAGTGCCTGGCGGCGGGTGACACAGGAGA | 2549 |
| Qy | 546 | AGTAGGCATATGTTATGAAGCTCTGTAGATGATGACGATTCAGATGGAGCTCAG | 605 |
| Db | 2548 | AGTAGGCATATGTTATGAAGCTCTGTAGATGATGACGATTCAGATGGAGCTCAG | 2489 |
| Qy | 606 | TATGCCAGAAAAATGGAGAAAAATACAACTGGGTGGACATTAACCAAGATTTTGA | 665 |
| Db | 2488 | TATGCCAGAAAAATGGAGAAAAATACAACTGGGTGGACATTAACCAAGATTTTGA | 2429 |
| Qy | 666 | AGAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAGCTATTTGGTCTTT | 725 |
| Db | 2428 | AGAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAGCTATTTGGTCTTT | 2369 |
| Qy | 726 | TGAAGCCATGCTGCTATTTGAATCATGATCCCAAGATGGATGCTGCATGATGGAAA | 785 |
| Db | 2368 | TGAAGCCATGCTGCTATTTGAATCATGATCCCAAGATGGATGCTGCATGATGGAAA | 2309 |
| Qy | 786 | CAAAGTTAATCGAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAT | 845 |
| Db | 2308 | CAAAGTTAATCGAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAT | 2249 |
| Qy | 846 | TAAAGATCTCACTTGCCTGAACCTAGCTAGGATTAAGATACATGTTTGTGCTTGTAT | 905 |
| Db | 2248 | TAAAGATCTCACTTGCCTGAACCTAGCTAGGATTAAGATACATGTTTGTGCTTGTAT | 2189 |
| Qy | 906 | AACGTGGTTAGAGGCCATTCACCTGGCACAGACAGTATTTACGTGCCCTTTACATTA | 965 |
| Db | 2188 | AACGTGGTTAGAGGCCATTCACCTGGCACAGACAGTATTTACGTGCCCTTTACATTA | 2129 |
| Qy | 966 | TCCAGACTTTTATGAGAGTCCCTGCTATGAAAGCTTTTCTCTGGAAATCTTGAATCTG | 1025 |
| Db | 2128 | TCCAGACTTTTATGAGAGTCCCTGCTATGAAAGCTTTTCTCTGGAAATCTTGAATCTG | 2069 |
| Qy | 1026 | TGACATTCGAAGGAAAAAGTAAATAAGCTGCTGTTTGAAGGAGGAGATTTTCAGTC | 1085 |
| Db | 2068 | TGACATTCGAAGGAAAAAGTAAATAAGCTGCTGTTTGAAGGAGGAGATTTTCAGTC | 2009 |
| Qy | 1086 | AATGACTTTATGATTTAAAAATGGCTTAACAGTGTGACAGATCTTCAGATTAACAGCATGCT | 1145 |
| Db | 2008 | AATGACTTTATGATTTAAAAATGGCTTAACAGTGTGACAGATCTTCAGATTAACAGCATGCT | 1949 |
| Qy | 1146 | AAAAGATGTGAGGATGATCATGCAAGAGAGTAAAGAGTACTCGAAGTGCACAGGAGA | 1205 |
| Db | 1948 | AAAAGATGTGAGGATGATCATGCAAGAGAGTAAAGAGTACTCGAAGTGCACAGGAGA | 1889 |
| Qy | 1206 | AGAAAGAGATCCAGAAGTTGAATAGAACCAACCAATGTTTGAAGTATTTACAGAGAT | 1265 |
| Db | 1888 | AGAAAGAGATCCAGAAGTTGAATAGAACCAACCAATGTTTGAAGTATTTACAGAGAT | 1829 |
| Qy | 1266 | GAAATTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAGAGACCAAGTC | 1325 |
| Db | 1828 | GAAATTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAGAGACCAAGTC | 1769 |
| Qy | 1326 | TGTTTCAGAGCTCAAAAATGATGGTTCAAGCAGAGATCTTCTTTCTGCAATTCATAA | 1385 |
| Db | 1768 | TGTTTCAGAGCTCAAAAATGATGGTTCAAGCAGAGATCTTCTTTCTGCAATTCATAA | 1709 |
| Qy | 1386 | TTCAATTGCATCATGCTCCAGGCCAGAAATGATACACAAAGAGAGATCATCCAAATAT | 1445 |

| | | | |
|----|------|----------------------------------------------------------------|------|
| Db | 1708 | TTCAATTGCATCATGGCATCCAGGCCAGAAATGATACTACAAAGAGATCATCCAATAT | 1649 |
| Qy | 1446 | GATGGGTTTTGAACCCCTTGTGAACAGAGAGCTACTTCCACTACCTTCCCTCGATATGC | 1505 |
| Db | 1648 | GATGGGTTTTGAACCCCTTGTGAACAGAGAGCTACTTCCACTACCTTCCCTCGATATGC | 1589 |
| Qy | 1506 | AAAAATTAATTAAGGGAAGAAATGGTGAACTATTTTTCAGAGATTAATAGATAGAAATAA | 1565 |
| Db | 1588 | AAAAATTAATTAAGGGAAGAAATGGTGAACTATTTTTCAGAGATTAATAGATAGAAATAA | 1529 |
| Qy | 1566 | AACCTGCTGTGAGGTTGTGAATTTAAACAAATTTACATTTGATTCCTGGATTTTTCTGTGA | 1625 |
| Db | 1528 | AACCTGCTGTGAGGTTGTGAATTTAAACAAATTTACATTTGATTCCTGGATTTTTCTGTGA | 1469 |
| Qy | 1626 | ATTTAGTGAAACAGTCACCATGCTGTTCTTCAAGATCTCTGTACAAACCACTTCTCTGT | 1685 |
| Db | 1468 | ATTTAGTGAAACAGTCACCATGCTGTTCTTCAAGATCTCTGTACAAACCACTTCTCTGT | 1409 |
| Qy | 1686 | GGATAACAAAAAGGCTTTTGGAACTCATCTCATGCAAGACATGGTGAAAGATGCACCTCG | 1745 |
| Db | 1408 | GGATAACAAAAAGGCTTTTGGAACTCATCTCATGCAAGACATGGTGAAAGATGCACCTCG | 1349 |
| Qy | 1746 | GTCTTTTGTGATCTCCGAGTGCTTCCGCCCAAGTGCTACCTATATAAATAATCACAG | 1805 |
| Db | 1348 | GTCTTTTGTGATCTCCGAGTGCTTCCGCCCAAGTGCTACCTATATAAATAATCACAG | 1291 |
| Qy | 1806 | GCTAAGGACTGTATGACTCTCTTGTACTCACTGTGTTCCGCCCACTCTGTAGTCTTAT | 1865 |
| Db | 1290 | GCTAAGGACTGTATGACTCTCTTGTACTCACTGTGTTCCGCCCACTCTGTAGTCTTAT | 1231 |
| Qy | 1866 | CAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTTGGTCATATTTCTGAGGAA | 1925 |
| Db | 1230 | CAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTTGGTCATATTTCTGAGGAA | 1171 |
| Qy | 1926 | TTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTTATAGGCAGAGAGGTTGATG | 1985 |
| Db | 1170 | TTTGCCACCTTGCAGGATG-----AGGCAGAGAGGTTGATG | 1134 |
| Qy | 1986 | CAGCGCTTACACCATGCTTTGAAACAGGAACCCCAAGGCAACATTTGGCTCTGTTAG | 2045 |
| Db | 1133 | CAGCGCTTACACCATGCTTTGAAACAGGAACCCCAAGGCAACATTTGGCTCTGTTAG | 1074 |
| Qy | 2046 | GTAACCTGGCTCTTTTACCATAACCTTCGCATTTATGATACAGTACCTTCTAAGTGCTTTG | 2105 |
| Db | 1073 | GTAACCTGGCTCTTTTACCATAACCTTCGCATTTATGATACAGTACCTTCTAAGTGCTTTG | 1014 |
| Qy | 2106 | AATTGGAACTCTACAGTATGACAGG--TACTATTTACATATATTTGATCTCTCTGAATTC | 2164 |
| Db | 1013 | AATTGGAACTCTACAGTATGACAGGATCTATTACATATATTTGATCTCTCTGAATTC | 954 |
| Qy | 2165 | CTTTACGCATGTTGATGTCAACATTTGAGTCTGCGGATGGCTCTCAATGGCAGAGAA | 2224 |
| Db | 953 | CTTTACGCATGTTGATGTCAACATTTGAGTCTGCGGATGGCTCTCAATGGCAGAGAA | 894 |
| Qy | 2225 | AGGATAATGGAAGACAGACAGAAAGCCGTAGTAGTAAAAAACAAGAAAAAAGAAA | 2284 |
| Db | 893 | AGGATAATGGAAGACAGACAGAAAGCCGTAGTAGTAAAAAACAAGAAAAAAGAAA | 834 |
| Qy | 2285 | GTTCCGCCATTTAGCCGAGAGATCACATGAGCCAGCATATCAGAACATGTGCTGGA | 2344 |
| Db | 833 | GTTCCGCCATTTAGCCGAGAGATCACATGAGCCAGCATATCAGAACATGTGCTGGA | 774 |
| Qy | 2345 | ATGTTTAAAAACATGGTAGCAATTTGACATGACCGCAAGTACGTTAAACCGAAGTTTGG | 2404 |
| Db | 773 | ATGTTTAAAAACATGGTAGCAATTTGACATGACCGCAAGTACGTTAAACCGAAGTTTGG | 714 |
| Qy | 2405 | CTTGATAGTGAACAGTTTCGGTATCAACACAGGTTTGTCTCCATTCACACAGTGTGATGACC | 2464 |
| Db | 713 | CTTGATAGTGAACAGTTTCGGTATCAACACAGGTTTGTCTCCATTCACACAGTGTGATGACC | 654 |
| Qy | 2465 | CCGCGCCAGTGCATCTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCT | 2524 |

Db 653 CCGCGCGCAGTGCACTTACAGTTCAAGGAAATGCTGACCTCAATAAATATAGCCCT 594
 Qy 2525 CCTCCTCAGTCTCCTGAACCTATGTGGCAGCTAGTAAGCACTTTCAACAGCGCAAAATG 2584
 Db 593 CCTCCTCAGTCTCCTGAACCTATGTGGCAGCTAGTAAGCACTTTCAACAGCGCAAAATG 534
 Qy 2585 ATATTGGAATAATCTCTAACCCGAGCATGAGGTTAATAGAAATTTAAAGGTTGCCAAA 2644
 Db 533 ATATTGGAATAATCTCTAACCCGAGCATGAGGTTAATAGAAATTTAAAGGTTGCCAAA 474
 Qy 2645 CCCAACTTTGGTTATGAGTTATGGCAGGAGCACAACAAAGGATCTAAAGTTCTT 2704
 Db 473 CCCAACTTTGGTTATGAGTTATGGCAGGAGCACAACAAAGGATCTAAAGTTCTT 414
 Qy 2705 CCTGAATTTGATTTCTCTGCTCATAAATATTTCTGTTGTGAAACTTGTTGAGAGAGA 2764
 Db 413 CCTGAATTTGATTTCTCTGCTCATAAATATTTCTGTTGTGAAACTTGTTGAGAGAGA 354
 Qy 2765 CTGGGAGGTGGCCATAAAGGGCAGAGCTCTTTTCAGACCAACTTTAGAGGGCACA 2824
 Db 353 CTGGGAGGTGGCCATAAAGGGCAGAGCTCTTTTCAGACCAACTTTAGAGGGCACA 294
 Qy 2825 TCACGAGCTCCACATCACCGGAAAGTGAGATGATTTCTTGGTAAACAACCTATTATAAG 2884
 Db 293 TCACGAGCTCCACATCACCGGAAAGTGAGATGATTTCTTGGTAAACAACCTATTATAAG 234
 Qy 2885 GAATACCTTTAGTTTGAAGCTTTATATGACATGAATGAAACTGCTGTTTAAAGTGGT 2944
 Db 233 GAATACCTTTAGTTTGAAGCTTTATATGACATGAATGAAACTGCTGTTTAAAGTGGT 174
 Qy 2945 TTATTATGTTCCATGGAGAACTGGTCTTATTGAATGCATGATGACGTTATATGTT 3004
 Db 173 TTATTATGTTCCATGGAGAACTGGTCTTATTGAATGCATGATGACGTTATATGTT 114
 Qy 3005 TTATTACAGATTTAATCACAAATCATTTTATGAATGATTGAGTGAATAAGTGTAT 3064
 Db 113 TTATTACAGATTTAATCACAAATCATTTTATGAATGATTGAGTGAATAAGTGTAT 54
 Qy 3065 AAGGTTAAATAATTTCTTGACAAAAA 3095
 Db 53 AAAGTTAAATAATTTCTTGACAAAAA 23

RESULT 3

ABV29823/c

ID ABV29823 standard; cDNA; 2668 BP.

XX AC ABV29823;

XX DT 16-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 29814.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX XX WO200160860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US05171.

XX PR 17-FEB-2000; 2000US-183319P.

XX PR 16-MAR-2000; 2000US-189862P.

XX PR 25-MAY-2000; 2000US-207454P.

XX PR 09-JUN-2000; 2000US-211314P.

XX PR 18-JUL-2000; 2000US-219007P.

XX PR 13-DEC-2000; 2000US-255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX XX

PI Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 6420; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate

CC cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer

CC in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a

CC patient;

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX SQ Sequence 2668 BP; 742 A; 617 C; 483 G; 820 T; 6 other;

Query Match 83.3%; Score 2577.8; DB 23; Length 2668;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 2638; Conservative 0; Mismatches 7; Indels 26; Gaps 3;

Qy 426 CGCATGCTGCACCTCCCGCTGGGCTGGGCTGAGAGGGAGGGGGCGGGCGGCCGA 485

Db 2668 CGCATGCTGCACCTCCCGCTGGGCTGGGCTGAGAGGGAGGGGGCGGGCGGCCGA 2609

Qy 486 GCGCGCTCGCTATTTCCTGGTCCGACAGTGGTGGCGCGGGTGACACCGGAGA 545

Db 2608 GCGCGCTCGCTATTTCCTGGTCCGACAGTGGTGGCGCGGGTGACACCGGAGA 2549

Qy 546 AGTAGGCATTAATGTTATGAAGCTTCTGTAGATGATGACGATTCAGGATGGGAGCTCAG 605

Db 2548 AGTAGGCATTAATGTTATGAAGCTTCTGTAGATGATGACGATTCAGGATGGGAGCTCAG 2489

Qy 606 TATGCCAGAAAATGAGAAAAGCAATACAACTGGGTGGACATTAACCAAGATTGGA 665

Db 2488 TATGCCAGAAAATGAGAAAAGCAATACAACTGGGTGGACATTAACCAAGATTGGA 2429

Qy 666 AGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTTCATGATAAGCTATTGCTCTTTT 725

Db 2428 AGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTTCATGATAAGCTATTGCTCTTTT 2369

Qy 726 TGAAGCCATGCTGCTATTGAAATGATGATGCCAAGATGCCATGCTGGCATGTTGAAA 785

Db 2368 TGAAGCCATGCTGCTATTGAAATGATGATGCCAAGATGCCATGCTGGCATGTTGAAA 2309

Qy 786 CCAAGTTAATCGAAAAGTTCTCAATTTTGAACAGCTATCAAGATGGCACTATTAAT 845

Db 2308 CCAAGTTAATCGAAAAGTTCTCAATTTTGAACAGCTATCAAGATGGCACTATTAAT 2249

Qy 846 TAAAGATCTCACCTTGGCTGAACTGATGAGGATTTATGATGATGATGATGATGATGAT 905

Db 2248 TAAAGATCTCACCTTGGCTGAACTGATGAGGATTTATGATGATGATGATGATGATGAT 2189

Qy 906 AACGTGGTTAGAAAGGCAATTCATGCGCACAGACAGTATTTACGTGCTTTTACATTTCA 965

Db 2188 AACGTGGTTAGAAAGGCAATTCATGCGCACAGACAGTATTTACGTGCTTTTACATTTCA 2129

Qy 966 TCCAGACTTTATAGAGATCCTGCTATGAAGGCTTTTGTCTCTGGAACTCTGAAAATCTG 1025

Db 2128 TCCAGACTTTATAGAGATCCTGCTATGAAGGCTTTTGTCTCTGGAACTCTGAAAATCTG 2069

Qy 1026 TGACATTGCAAGGAAAAAGTAATAAGTCTGCTGTTTTTTTGAAGAGGAGATTTTCAGTC 1085

Db 2068 TGACATTTGCAAGGAAAGAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTC 2009
Qy 1086 AATGACATATGGAATTTAAATGGCTTAACAGTGTGACAGATCTTCCAGATTTACAGCATGCT 1145
Db 2008 AATGACATATGGAATTTAAATGGCTTAACAGTGTGACAGATCTTCCAGATTTACAGCATGCT 1949
Qy 1146 AAAAGATGTGGAGTGACATGCAAGAAAGAGTAAAGAGTACTCGAAGTGCACAAAGGAGA 1205
Db 1948 AAAAGATGTGGAGTGACATGCAAGAAAGAGTAAAGAGTACTCGAAGTGCACAAAGGAGA 1889
Qy 1206 AGAAAGAGATCCAGAGTGTGAATAGAACACCAACCAATGTTTAGCAGATTTTCAGCAGAGT 1265
Db 1888 AGAAAGAGATCCAGAGTGTGAATAGAACACCAACCAATGTTTAGCAGATTTTCAGCAGAGT 1829
Qy 1266 GAATATTTACTCGTGTGTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAGAGACCAAGTGC 1325
Db 1828 GAATATTTACTCGTGTGTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAGAGACCAAGTGC 1769
Qy 1326 TGTTCGCAAGACTCAAAAATTTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCATAA 1385
Db 1768 TGTTCGCAAGACTCAAAAATTTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCATAA 1709
Qy 1386 TTCAATTCGATCATGGCATCCAGGCCAGAAATGATACTACAAAGAGATCATCCAAATTTAT 1445
Db 1708 TTCAATTCGATCATGGCATCCAGGCCAGAAATGATACTACAAAGAGATCATCCAAATTTAT 1649
Qy 1446 GATGGTGTTCGAAACCCCTGTGTAACCCAGAGGCTACTTCCACCTACTCTCCCTCGATATGC 1505
Db 1648 GATGGTGTTCGAAACCCCTGTGTAACCCAGAGGCTACTTCCACCTACTCTCCCTCGATATGC 1589
Qy 1506 AAAAATAAATTAAGGGAAGAAATGGTGAACTATTTTCAAGATTAATAGATAGAAATAA 1565
Db 1588 AAAAATAAATTAAGGGAAGAAATGGTGAACTATTTTCAAGATTAATAGATAGAAATAA 1529
Qy 1566 AACTGTCTGTGAGGTGTGAATTTAACTAATTTACATTTGATCTCTGGATTTTCTGTGA 1625
Db 1528 AACTGTCTGTGAGGTGTGAATTTAACTAATTTACATTTGATCTCTGGATTTTCTGTGA 1469
Qy 1626 ATTTAGTGAAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTTACAAACCACTTTCCCTGCT 1685
Db 1468 ATTTAGTGAAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTTACAAACCACTTTCCCTGCT 1409
Qy 1686 GGATAACAAAAGGCTTTTGGAACTCATCTCATGCAAGACATGTTGAAAGATGCATTCG 1745
Db 1408 GGATAACAAAAGGCTTTTGGAACTCATCTCATGCAAGACATGTTGAAAGATGCATTCG 1349
Qy 1746 GTCTTTTCTCAGATCTCCGAGTGTCTTCCCCCAAGTGTCTACTATATAATAATCACCAG 1805
Db 1348 GTCTTTTCTCAG - -TCTCTCGGTGCTTTCCCCCAAGTGTCTACTATATAATAATCACCAG 1291
Qy 1806 GCTAAGGACTGTATCGACTCTTTGTTTACTCACTGTGTGCGCCATTTCTGTAGTCTTAT 1865
Db 1290 GCTAAGGACTGTATCGACTCTTTGTTTACTCACTGTGTGCGCCATTTCTGTAGTCTTAT 1231
Qy 1866 CAGATCCATGGACATAACAGGCTCGACAGAGATTAAGCTTGGTCATATTTCTTGAGGAA 1925
Db 1230 CAGATCCATGGACATAACAGGCTCGACAGAGATTAAGCTTGGTCATATTTCTTGAGGAA 1171
Qy 1926 TTTTGCCACTTCAGAGTGTGATTTATGACATTTTATTTTAAAGCAGAGAGGTTGTATG 1985
Db 1170 TTTTGCCACTTCAGAGTGTGATTTTAAAGCAGAGAGGTTGTATG 1134
Qy 1986 CAGCGCTTCACACCATGTGTTGAAACAGGAACCCCAAGGCAACATTTGGCTGTTTAG 2045
Db 1133 CAGCGCTTCACACCATGTGTTGAAACAGGAACCCCAAGGCAACATTTGGCTGTTTAG 1074
Qy 2046 GTACCTGGTCTCTTTACCATACCTTCGCAATTTATGATACAGTACCTTCTAAGTGGCTTGG 2105
Db 1073 GTACCTGGTCTCTTTACCATACCTTCGCAATTTATGATACAGTACCTTCTAAGTGGCTTGG 1014
Qy 2106 AATTGGAACTCTACAGTATGACAGAG -TACTATTACATATATTTGTTATCTCTCAATTC 2164

Db 1013 AATTGGAACCTCAGTATGACAGAGATACATTATACATATATTGGTATCTCTCTGAATTC 954
Qy 2165 CTTTACCACATGGTTGATGTCAACATTTGAGTCGTCGCCATGGCTCTCAAAATGGCAGAGGAA 2224
Db 953 CTTTACGATGGTTGATGTCAACATTTGAGTCGTCGCCATGGCTCTCAAAATGGCAGAGGAA 894
Qy 2225 AGGATAATGGAAGAGCAGCAGAAAGCGCTGATGTAGTATTAATAAACAAGAAAGAAAGAA 2284
Db 893 AGGATAATGGAAGAGCAGCAGAAAGCGCTGATGTAGTATTAATAAACAAGAAAGAAAGAA 834
Qy 2285 GTTCGCCCATTTGAGCCGAGAGATCAAAATGAGCAAGCATATCAGAACATGTGTCTGGA 2344
Db 833 GTTCGCCCATTTGAGCCGAGAGATCAAAATGAGCAAGCATATCAGAACATGTGTCTGGA 774
Qy 2345 ATGTTTAAAACCATGGTAGCATTTGACATGACCGGCAAGTACGTAAACCGAAAGTTTGG 2404
Db 773 ATGTTTAAAACCATGGTAGCATTTGACATGACCGGCAAGTACGTAAACCGAAAGTTTGG 714
Qy 2405 CTTGATAGTGAAACAAAGTTCGGTATGAACACACAGGTTTGTCTCATTCACACAGTGTGATGACC 2464
Db 713 CTTGATAGTGAAACAAAGTTCGGTATGAACACACAGGTTTGTCTCATTCACACAGTGTGATGACC 654
Qy 2465 CCGCGCCAGTGCATCTTACAGTTCGAAGGAAATGTCTGACCTCAATAAATATAGCCCT 2524
Db 653 CCGCGCCAGTGCATCTTACAGTTCGAAGGAAATGTCTGACCTCAATAAATATAGCCCT 594
Qy 2525 CTTCTCAGTCTCTGAACTGTATGTGGCAGCTAGTAGACACTTTTCAACAGCAGAAATG 2584
Db 593 CTTCTCAGTCTCTGAACTGTATGTGGCAGCTAGTAGACACTTTTCAACAGCAGAAATG 534
Qy 2585 ATATTGGAATAATTTCTTAACCCGACCATGAGGTTTATAGAAATTTTAAAGTTTCCCAA 2644
Db 533 ATATTGGAATAATTTCTTAACCCGACCATGAGGTTTATAGAAATTTTAAAGTTTCCCAA 474
Qy 2645 CCCAACTTTGGTTTATGAAGTTATGCGCAGCAGACACAAAGGAATCTTAAAGTTCT 2704
Db 473 CCCAACTTTGGTTTATGAAGTTATGCGCAGCAGACACAAAGGAATCTTAAAGTTCT 414
Qy 2705 CCTGAAATTTGATTTCTCTGCTCATAAATATTTTCTGTTGTGAAACTTTGTTGAGAGAGA 2764
Db 413 CCTGAAATTTGATTTCTCTGCTCATAAATATTTTCTGTTGTGAAACTTTGTTGAGAGAGA 354
Qy 2765 CTGGGAGGTCGCGCATTAAGGGGAGAGTCTTCTTTTCAGACCTTTCAGAGGCGACA 2824
Db 353 CTGGGAGGTCGCGCATTAAGGGGAGAGTCTTCTTTTCAGACCTTTCAGAGGCGACA 294
Qy 2825 TCACCAGGCTCCACATCACGGGAGTGAATGATTTCTTGGGTAAACAACCTCATTATAAG 2884
Db 293 TCACCAGGCTCCACATCACGGGAGTGAATGATTTCTTGGGTAAACAACCTCATTATAAG 234
Qy 2885 GAATACTTTTGTAGTTTGACAGCTTATATGACATGAATGAAACTGCTGTTTAAAGTGT 2944
Db 233 GAATACTTTTGTAGTTTGACAGCTTATATGACATGAATGAAACTGCTGTTTAAAGTGT 174
Qy 2945 TTATTTATTTCCATGGAGAACTGGTCTTATTTGATGATGATGATGATGATGATGATGAT 3004
Db 173 TTATTTATTTCCATGGAGAACTGGTCTTATTTGATGATGATGATGATGATGATGATGAT 114
Qy 3005 TTATTTACAGATTTAATCAAAATCAATTTTTTATGATGATGATGATGATGATGATGATGAT 3064
Db 113 TTATTTACAGATTTAATCAAAATCAATTTTTTATGATGATGATGATGATGATGATGATGAT 54
Qy 3065 AAAGGTTTAAATTTCTTGACAAAAA 3095
Db 53 AAAGGTTTAAATTTCTTGACAAAAA 23

RESULT 4
AAL49929
ID AAL49929 standard; cDNA; 2488 BP.
XX
AC AAL49929;
XX

Db 1261 AGATGCACTTCGGTCTTTTGTGTCAGTCTCTCCG--GTGCTTTTCCCCCAAGTGCTACCTATAT 1318
QY 1794 AATAATCACCAGCTAAGAGCTGTATCGACTCCTCTTTGTTACTCACCTGTGTTCGGCCATTC 1853
Db 1319 AATAATCACCAGCTAAGAGCTGTATCGACTCCTCTTTGTTACTCACCTGTGTTCGGCCATTC 1378
QY 1854 TGTAGTCTTTATTCAGATTCATGACATACAGGGCTCGACAGAGATAGCTTGTGTCAT 1913
Db 1379 TGTAGTCTTTATTCAGATTCATGACATACAGGGCTCGACAGAGATAGCTTGTGTCAT 1438
QY 1914 ATTCTTGAGGAATTTCCCACTTTCAGGATGAGTTTATGACATTTTATTTTATATGAGCAG 1973
Db 1439 ATCTTGAGGAATTTCCCACTTTCAGGATG-----AGGCAG 1475
QY 1974 AGAAGTTGATCAGCGCTTCACCACTGCTGTGTGAACAGGAAACCCCAAGGCAACATT 2033
Db 1476 AGAAGTTGATCAGCGCTTCACCACTGCTGTGTGAACAGGAAACCCCAAGGCAACATT 1535
QY 2034 TGGCCTGTTTAGTACTCGGTCTTTTACATTAACCTTCGCATTATGATACAGTACTTTC 2093
Db 1536 TGGCCTGTTTAGTACTCGGTCTTTTACATTAACCTTCGCATTATGATACAGTACTTTC 1595
QY 2094 TAAGTGGCTTTGAATTTGGAATCTACAGTATGACGAGTACTATTACATATATTGTTATC 2153
Db 1596 TAAGTGGCTTTGAATTTGGAATCTACAGTATGACGAGTACTATTACATATATTGTTATC 1655
QY 2154 TCTCTGAATTCCTTTAGCATGTTGATGTCATCATGTCGTCGGATGCTCTCAAA 2213
Db 1656 TCTCTGAATTCCTTTAGCATGTTGATGTCATCATGTCGTCGGATGCTCTCAAA 1715
QY 2214 TGGCAGGAAAGGATAATGGAAGACAGCAGAAAGCCGTAGTGTAAAGAAACAAAG 2273
Db 1716 TGGCAGGAAAGGATAATGGAAGACAGCAGAAAGCCGTAGTGTAAAGAAACAAAG 1775
QY 2274 AAAAAAGAAAGTTGCCCATTTAGCCGAGAGATCAATGAGCCCAAGCATATCAAGACA 2333
Db 1776 AAAAAAGAAAGTTGCCCATTTAGCCGAGAGATCAATGAGCCCAAGCATATCAAGACA 1835
QY 2334 TGTGTGCTGGAATGTTTAAACCATGTTAGCATTTGACATGAGCGGCAAGTACGTTAAC 2393
Db 1836 TGTGTGCTGGAATGTTTAAACCATGTTAGCATTTGACATGAGCGGCAAGTACGTTAAC 1895
QY 2394 CGAAGTTTGAGCTTGATAGTGAACAAGTTCCGATGACACACAGGTTTGTCTCAATCAACA 2453
Db 1896 CGAAGTTTGAGCTTGATAGTGAACAAGTTCCGATGACACACAGGTTTGTCTCAATCAACA 1955
QY 2454 GTGTGATGACCCCGCCGCGCAGTGCATCTTACAGTTCGAAGAAATGTCGACCTCAATA 2513
Db 1956 GTGTGATGACCCCGCCGCGCAGTGCATCTTACAGTTCGAAGAAATGTCGACCTCAATA 2015
QY 2514 AATATAGCCCTCCTCTCAGTCTCCTGAACTGTATGTCGACGCTAGTACGACTTTCAAC 2573
Db 2016 AATATAGCCCTCCTCTCAGTCTCCTGAACTGTATGTCGACGCTAGTACGACTTTCAAC 2075
QY 2574 AGGCAAAATATATTTGAAAAATATTCCTAACCCGACCATGAGGTTAATAGAAATTTTAA 2633
Db 2076 AGGCAAAATATATTTGAAAAATATTCCTAACCCGACCATGAGGTTAATAGAAATTTTAA 2135
QY 2634 AGGTTCCAAAACCCAACTTTGTGTTATGAAGTTATGCGAGGAGACACAAAAGGAAT 2693
Db 2136 AGGTTCCAAAACCCAACTTTGTGTTATGAAGTTATGCGAGGAGACACAAAAGGAAT 2195
QY 2694 CTAAAGTTCTCTGAAATTTGATTTCTCTGCTCATTAATTTTCTGTTGTGAACCTTG 2753
Db 2196 CTAAAGTTCTCTGAAATTTGATTTCTCTGCTCATTAATTTTCTGTTGTGAACCTTG 2255
QY 2754 TTTGAGAGACTGCGGAGGTGSCCAATAAGGGCGAGAGTCTTCTTTTCAGACCCAACTCT 2813
Db 2256 TTTGAGAGACTGCGGAGGTGSCCAATAAGGGCGAGAGTCTTCTTTTCAGACCCAACTCT 2315
QY 2814 TAGAGGGCATCATCAGGCTCCACATCAGGGAACTGAGTGTCTTGGTAAACAA 2873
Db 2316 TAGAGGGCATCATCAGGCTCCACATCAGGGAACTGAGTGTCTTGGTAAACAA 2375

QY 2874 CTCATTATAAGGAATACATTTTAGTTTGACAGCCTTATATGACATGAATGAAACTGCTGT 2933
Db 2376 CTCATTATAAGGAATACATTTTAGTTTGACAGCCTTATATGACATGAATGAAACTGCTGT 2435
QY 2934 TTTAAAGTGGTTTATATATGTTTCCATGGAAGAAACTGGTCTTATTGAAT 2981
Db 2436 TTTAAAGTGGTTTATATATGTTTCCATGTAAGACACTGGGTTCATTAAT 2483
RESULT 5
ABS51309
ID ABS51309 standard; cDNA; 719 BP.
XX
AC ABS51309;
XX
DT 21-OCT-2002 (first entry)
XX
DE cDNA encoding human secretory protein #7.
XX
KW Human; secretory polypeptide; SPTM; actinic keratosis; arteriosclerosis;
KW bursitis; cirrhosis; hepatitis; polycythemia vera; anaemia; psoriasis;
KW primary thrombocytopenia; cancer; adenocarcinoma; leukaemia; myeloma;
KW sarcoma; immune system disorder; acquired immunodeficiency syndrome;
KW AIDS; allergy; aschma; Crohn's disease; diabetes mellitus; gout;
KW glomerulonephritis; Goodpasture's syndrome; thyroiditis; pancreatitis;
KW hepatitis; multiple sclerosis; osteoporosis; Reiter's syndrome;
KW rheumatoid arthritis; neurological disorder; epilepsy; stroke; dementia;
KW Alzheimer's disease; Pick's disease; Huntington's disease; mood; anxiety;
KW Parkinson's disease; central nervous system disorder; mental disorder;
KW schizophrenic disorder; amnesia; Tourette's disorder; transgenic animal;
KW gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200257304-A2.
XX
PD 25-JUL-2002.
XX
PF 15-JAN-2002; 2002WO-US01340.
XX
PR 16-JAN-2001; 2001US-261864P.
PR 16-JAN-2001; 2001US-261865P.
PR 16-JAN-2001; 2001US-261979P.
PR 16-JAN-2001; 2001US-261981P.
PR 17-JAN-2001; 2001US-262164P.
PR 17-JAN-2001; 2001US-262208P.
PR 17-JAN-2001; 2001US-263131P.
PR 19-JAN-2001; 2001US-262599P.
PR 19-JAN-2001; 2001US-262760P.
PR 19-JAN-2001; 2001US-263063P.
PR 19-JAN-2001; 2001US-263066P.
PR 19-JAN-2001; 2001US-263069P.
PR 19-JAN-2001; 2001US-263070P.
PR 19-JAN-2001; 2001US-263074P.
PR 19-JAN-2001; 2001US-263076P.
PR 19-JAN-2001; 2001US-263077P.
PR 19-JAN-2001; 2001US-263329P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Panzer SR, Lincoln SE, Altus CM, Dufour GE, Hillman JL, Jones AL;
PI Dam TC, Liu TF, Harris B, Flores V, Daffo A, Marwaha R, Chen AJ;
PI Chang SC, Gerstin EH, Peralta CH, David MH, Lewis SA;
XX
WPI: 2002-590716/63.
DR P-PSDB; ABG69816.
XX
PT New purified secretory polypeptides and polynucleotides, useful in the
PT diagnosis, study, prevention or treatment of diseases associated with
PT decreased expression of functional secretory molecules, e.g. AIDS,
PT cancer or allergies -


```
CC seqdata.uspto.gov/sequence.html?DocID=20020137160.
XX
SQ Sequence 398 BP; 131 A; 89 C; 85 G; 93 T; 0 other;

  Query Match      11.3%; Score 350.8; DB 25; Length 398;
  Best Local Similarity 94.3%; Pred. No. 2.1e-77;
  Matches 364; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 2298 GCCGAGAGATCACATGAGCCCAAGCATATCAGAACATGTGCTGGAATGTTTAAACCA 2357
Db 13 GCCGAGAGATCACATGAGCCCAAGCATATCAGAACATGTGCTGGAATGTTTAAACCA 72

Qy 2358 TGGTAGCATTTGACATGAGCCCAAGTACGTAAACCGAAGTTTGAGCTTGATGTAAC 2417
Db 73 TGGGAGCTTTTGACATGGATGCGCAAGTACGAAACCCCAAGTTTGAGCTTGATGTAAC 132

Qy 2418 AAGTTCGGTATGACACACAGGTTTGCTCCATTCAACAGGTGTGATGACCCGCGCCAGTGC 2477
Db 133 AAGTTCGGTATGACACACAGATTGCTCCATTCAACAGGTGTGATGACACACACAGTGC 192

Qy 2478 ACTACTTACAGTTCAAGGAATGTCTGACCTCAATPAATATATAGCCCTCTCTCTAGTCTC 2537
Db 193 ACTATCTGAGTCAAGGAATGTCTGACCTCAATPAATATATAGCCCTCTCTCTAGTCTC 252

Qy 2538 CTGAACCTGATGTGCAGCTAGTAAGCACTTTCAACAGGCAAAATGATTTGAAATA 2597
Db 253 CAGAACTGATGTGCAGCTAGTAAGCACTTTCAACAGGCAAAATGATTTGAAATA 312

Qy 2598 TTCTTAACCCGACCATGAGGTTAATAGAAATTTTAAAGTTGCCAAACCCCAACTTTGTGG 2657
Db 313 TCCCAAAACCCAGACCATGAGGTTAATAGAAATTTTAAAGTTGCCAAACCCCAACTTTGTGG 372

Qy 2658 TTATGAAGTTATTGGCAGGAGACAC 2683
Db 373 TTATGAAGTTATTGGCAGGAGACAC 398
```

```
RESULT 7
ABQ51658
ID ABQ51658 standard; DNA; 799 BP.
XX
AC ABQ51658;
XX
XX
DT 12-JUL-2002 (first entry)
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 38249.
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
XX WO200218632-A2.
XX
PD 07-MAR-2002.
XX
XX 01-SEP-2001; 2001WO-EP10074.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX 05-SEP-2000; 2000DE-1044543.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX PT amplicons from chemically treated DNA
```

```
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 799 BP; 118 A; 95 C; 297 G; 289 T; 0 other;
```

```
  Query Match      11.1%; Score 343.8; DB 24; Length 799;
  Best Local Similarity 83.5%; Pred. No. 1.7e-75;
  Matches 390; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 87 TGCTCGTTTCGTCCGCAACCACTAAGCTCTACGCAAACTCCACGGTTTCCTTCGGCT 146
Db 2 TGTCGTTTCGTTCGTAATATTAAAGGTTTACGTAATTTTACGGTTTTTTTCGTTT 61

Qy 147 TCGGCTCACCTTTCTAAGAAATTCAGAGGCGCAGCAGCGGGCTCTGAGACT 206
Db 62 TCGGTTATTTTAAAGAAATTTTACGGGTAGCTAGACGGGCGGGTTTCGATT 121

Qy 207 CCGGCTCCGCTCTTTCCGGAAACCGCCCACTACCCAGGACTCCGACAGGGGTGAAA 266
Db 122 TCGGTTTCGTTTTTTTTCGGGAATCGTTTATTATTATTAGGATTTTCGATAGGGTGAAA 181

Qy 267 AAGATACTTCGGTCTCGCATCGTCTCTAATCTCGCAGAGAGAGAGGGCGCCCAT 326
Db 182 AAGATAATTTTCGGTTTCGCGATCGTTTTTAATTTTCGCGAAGAGAGGGCGTCTAT 241

Qy 327 CGGCGAAACGAGGCGGTGCGAGGAGGGGTGTGGCGGGAGCGCGAAGTCCCGGG 386
Db 242 CGGTGCAACGAGGCGGTGCGAGGAGGGGTGTGGTTCGGGAGCGCGAAGTTTCGG 301

Qy 387 AGTAAGGAGAGGGGCGGGGTGCGGCGTCCCGGCATACCGCATCGCTGCACGTCGG 446
Db 302 AGTAAGGAGAGGGGCGGGGTGCGGCGTTCGGGTATACGTTATCGTGTACGTTGTCG 361

Qy 447 TCGGCTCGGCTGAGAGGGAGGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 506
Db 362 TCGGTTGCGTTGAGAGGGGAGGGGCGGCGCGCGCGCGCGCGCGCGCGCGCG 421

Qy 507 GTCCGACAGTCCGTGGCGCGCGCGGGTGACCAACCGGAGAGTAGCA 553
Db 422 GTTCGATAGTCCGTGGCGCGCGGGTGATTACGCGAGAGTAGTA 468
```

```
RESULT 8
ABQ51659/c
ID ABQ51659 standard; DNA; 799 BP.
XX
AC ABQ51659;
XX
XX
DT 12-JUL-2002 (first entry)
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 38250.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
```

KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 XX WO200218632-A2.
 XX
 XX PD 07-MAR-2002.
 XX
 XX PF 01-SEP-2001; 2001WO-EP10074.
 XX
 XX PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX PR 05-SEP-2000; 2000DE-1044543.
 XX
 XX PA (EPIC-) EPIGENOMICS AG.
 XX
 XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX
 XX DR WPI; 2002-371829/40.
 XX
 XX PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA -
 XX
 XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.
 XX
 XX SQ Sequence 799 BP; 289 A; 297 C; 95 G; 118 T; 0 other;
 Query Match 11.1%; Score 343.8; DB 24; Length 799;
 Best Local Similarity 83.5%; Pred. No. 1.7e-75;
 Matches 390; Conservative 0; Mismatches .77; Indels 0; Gaps 0;

QY 87 TGCTCGTTCGTCGCAACCACTAAGTCTACGCAACCTCCACGGTTCCTTCCGCT 146
 DB 798 TGTTCGTTCTGTTTCGTAATTAATTAAGGTTTACGTAATTTTACGGTTTTTTCGTT 739
 QY 147 TCGCGTCACTTCTAAGAAATCCACAGGGACGACGCGGGGCTCTGAGACT 206
 DB 738 TCGGTTATTTTAAAGAAATTTTAAAGGTTAGCGTAGACGGGGCGGTTTGGAGTT 679
 QY 207 CCGCGCTCCGCTCTTTCCGGGAACCCGCACTACCAGGACTCCGACAGAGGTGAAA 266
 DB 678 TCGGTTTCGTTTTCGGAATCGTTTATTTAGGATTCGATAGGGGTGAAA 619
 QY 267 AAGTAATCTTCGTCGTCGATCGTCTTAATCTCGGAGAAAGAGAGCGCGCCAT 326
 DB 618 AAGATAATTTTCGGTTTCGCGATCGTTTAAATTCGCGAAGAGAGCGGTCGTTAT 559
 QY 327 CGGCGGAACGAGCGGTGGCGAGGGGGTGTGCGCGGGAGCGGAGTCCCGGG 386
 DB 558 CGGTCGAACGAGCGGTGGCGAGGGGGTGTGTCGGGGAGCGCGAAGTTTTCGG 499

QY 387 AGTAAGGAGAGGGGGCGGCTCCGGCGTCCCGGCGATACGATCGTCACGCTCCCG 446
 DB 498 AGTAAGGAGAGGGGGCGGCTCCGGCGTTCGGGTATACGATCGGTGACGTTGTCGG 439
 QY 447 TCGGCTCGGCTGAGAGGGGAGGGGGCGGCGCGCCGAGGCGCGTCTGTTATTTCCG 506
 DB 438 TCGGTTTCGGTTGAGAGGGGAGGGGGCGGCGCGTCCGAGGCGGTCGTTATTTCCG 379
 QY 507 GTCCGACAGTGCCTGCGCGCGCGCGTGCACACCGGAGAGTAGCA 553
 DB 378 GTTCGATAGTCCGTGCGCGCGCGGTGATTACGGGAGAGTAGTA 332

RESULT 9

ABV15331/c
 ID ABV15331 standard; cDNA; 356 BP.

XX AC ABV15331;

XX DT 13-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 15322.

XX KW Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX FN WO200160860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US05171.

XX PR 17-FEB-2000; 2000US-183319P.

XX PR 16-MAR-2000; 2000US-189862P.

XX PR 25-MAY-2000; 2000US-207454P.

XX PR 09-JUN-2000; 2000US-211314P.

XX PR 18-JUL-2000; 2000US-219007P.

XX PR 13-DEC-2000; 2000US-255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Endege WO, Monahan JE;

XX DR WPI; 2001-662795/76.

XX PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -

XX FS Claim 1; Page 2572; 11750pp; English.

XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV6213) of the
 CC specification or its complement. (I) is useful for:

XX CC (a) assessing whether a patient is afflicted with prostate cancer;

XX CC (b) monitoring the progression of prostate cancer in a patient;

XX CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;

XX CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;

XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;

XX CC (f) assessing the prostate cell carcinogenic potential of a compound;

XX CC (g) determining whether prostate cancer has metastasized in a patient;

XX CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;

XX CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX SQ Sequence 356 BP; 97 A; 80 C; 85 G; 94 T; 0 other;

Query Match 9.4%; Score 292.4; DB 23; Length 356;

```
Best Local Similarity 90.5%; Pred. No. 7.7e-63;
Matches 344; Conservative 0; Mismatches 11; Indels 25; Gaps 2;

QY 1671 AACCACTTCTCTGGTGGATACAAAGGCTTTTGGAACTCATCTCATGCAAGACATGGT 1730
   |||
Db 355 AAACCACTTTCGGTGGATACAAAGGCTTTTGGAACTCATCTCATGCAAGACATGGT 296

QY 1731 GAAAGATCACTCGGTCTTTTGTGAGATCCCTCCGAGTGCCTTCCCAAGTGTACCTA 1790
   |||
Db 295 GAAAGATCACTCGGTCTTTTGTGAGATCCCTCCGAGTGCCTTCCCAAGTGTACCTA 238

QY 1791 TATAATATCAACAGGCTAAGACTGTATCGACTCTTTGTACTCACTGTGTTCGGCCA 1850
   |||
Db 237 TATAATATCAACAGGCTAAGACTGTATCGACTCTTTGTACTCACTGTGTTCGGCCA 178

QY 1851 TTCTGTAGTCTTATTTCAGATCCATGACATACAGGGCTTCGACAGAGAGATAAGCTTGGT 1910
   |||
Db 177 TTCTGTAGTCTTATTTCAGATCCATGACATACAGGGCTTCGACAGAGAGATAAGCTTGGT 118

QY 1911 CATATCTTGGAGAAATTTGCCACCTTGCAGGTAGATGATTTATGACATTTTATTAATAGG 1970
   |||
Db 117 CATATCTTGGAGAAATTTGCCACCTTGCAGGTAGATGATTTATGACATTTTATTAATAGG 81

QY 1971 CAGAGAAGGTTGATCAGCGCTTACACCATCTGTGAAACAGGAACCCCAAGGCAAC 2030
   |||
Db 80 CAGAGAAGGTTGATCAGCGCTTACACCATCTGTGAAACAGGAACCCCAAGGCAAC 21

QY 2031 ATTTGGCCCTGTTAGGTACC 2050
   |||
Db 20 ATTTGGCCCTGTTAGGTACC 1

RESULT 10
ABV36125/c
ID ABV36125 standard; cDNA; 439 BP.
XX
AC ABV36125;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 36116.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
XX
PR 16-MAR-2000; 2000US-189862P.
XX
PR 25-MAY-2000; 2000US-207454P.
XX
PR 09-JUN-2000; 2000US-211314P.
XX
PR 18-JUL-2000; 2000US-219007P.
XX
PR 13-DEC-2000; 2000US-255281P.
XX
(PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
XX
DR Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim.1; Page 7483; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
```

```
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 439 BP; 113 A; 104 C; 110 G; 112 T; 0 other;
   |||
Query Match 9.3%; Score 288.8; DB 23; Length 439;
Best Local Similarity 91.4%; Pred. No. 6.8e-62;
Matches 338; Conservative 0; Mismatches 7; Indels 25; Gaps 2;

QY 1687 GATAACAAAAGGCTTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACCTCGG 1746
   |||
Db 439 GATAACAAAAGGCTTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACCTCGG 380

QY 1747 TCTTTTGTGAGATCCCTCGAGTGTCTTCCCAAGTGTACCTATATATATATCATCCAGG 1806
   |||
Db 379 TCTTTTGTGAG--TCCCTCCGGTGTCTTCCCAAGTGTACCTATATATATATCATCCAGG 322

QY 1807 CTAAAGGACTGTATCGACTCTTGTACTCTGTTGGCCATCTCTAGTCTTATTC 1866
   |||
Db 321 CTAAAGGACTGTATCGACTCTTGTACTCTGTTGGCCATCTCTAGTCTTATTC 262

QY 1867 AGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTTCATATTTCTTGAGGAAT 1926
   |||
Db 261 AGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTTCATATTTCTTGAGGAAT 202

QY 1927 TTGCCACCTTGCAGATGAGTTTATGACATTTTATTTAATAGCAGAGAAGTTGATGC 1986
   |||
Db 201 TTGCCACCTTGCAGGATG-----AGGCAGAGAGAAGTTGATGC 165

QY 1987 AGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAGG 2046
   |||
Db 164 AGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAGG 105

QY 2047 TACCTGGGTC 2056
   |||
Db 104 TACCTCGGCC 95

RESULT 11
ABV45182/c
ID ABV45182 standard; cDNA; 439 BP.
XX
AC ABV45182;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 45173.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
XX
PR 16-MAR-2000; 2000US-189862P.
```



```

Db 619 AAAAAAATACCTCCGATCTCGCATCGTCTCTAATCTCGGAAAAAATAAAGACGACGCCA 560
Qy 326 TCGSCCGAACGAGCGCGGTGCGGAGGAGGCGGTGCGCGGAGCGCGAAGTCCCGCG 385
Db 559 TCGACCGAAGCAAAACGATACGAAAAAATAAATACCGAAAAACGCGAAATCCCGGA 500
Qy 386 GAGTAAGGAGAGGGGCGCGGTGCGCGTCCCGGCGATACGATGCGTGCACGCTGCCG 445
Db 499 AATATAAAAAAATAAATCGCGTCCCGAATACGATACGATACGATACGATACGATACG 440
Qy 446 GTCCGGCTGGCTGAGAGGGAGGGGCGCGGCGCGGCGCGGCGCGGCGCGTTCGTTTCG 505
Db 439 ATCGAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 380
Qy 506 GGTCCGACGACGTGCGTGGCGCGCGGTGACACCGGAGAGAGTAGGCATAA 556
Db 379 AATCCGAACAATACGTAACGACGCGAATAACACGAAAAAATAAATAAATAA 329

```

RESULT 13

ABQ51661
ID ABQ51661 standard; DNA; 799 BP.

AC ABQ51661;

XX 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 38252.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPT; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue

CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX Sequence 799 BP; 320 A; 262 C; 95 G; 122 T; 0 other;

Query Match 8.9%; Score 274.2; DB 24; Length 799;

Best Local Similarity 73.9%; Pred. No. 4.1e-58;
Matches 348; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 86 CTGCTCGTTCGTGTCGCAACCACTAAGTCTACGAAAACCTCCACGTTTCTCTTCGCC 145

Db 1 CTACTCGTTGCTACTCGCAACCACTAATAATCTAGCAAACTCCACGATTTCTTTCGCC 60

Qy 146 TTCGCGTCACTTCTTAAGAAATTCGAGAGGCGAGCGAGCGGGGGGGCTCTGAGAC 205

Db 61 TTCGCGTCACTTCTTAAGAAATTCGAGAGGCGAGCGAGCGGGGGGGCTCTAATAAC 120

Qy 206 TCCGGGCTCCGCTCTTTTCCGGGAACCGCCCACTACCCAGGACTCCGACAGAGGTGAAA 265

Db 121 TCCGAATCCGCTCTTTTCCGAAAACCGCCCACTACCCAAAATCCGACAAAAATAAA 180

Qy 266 AAGATTAATTCGCGTCTCGCGATCGTCTTAATCTCGCGAGAGAGAGCGCGCGCCA 325

Db 181 AAAAAATACTTCCGATCTCGCGATCGTCTTAATCTCGCGAAAAAAGAAAAACGACCGCA 240

Qy 326 TCGSCCGAACGAGCGGTGCGGAGGAGGGGTGCGCGGGGAGCGGAGTCCCGCG 385

Db 241 TCGACCGAACGAAACGATACGAAAAAATAAATAATACCGAAAAACCGAAATCCCCGA 300

Qy 386 GAGTAAGGAGAGGGGCGCGGTGCGCGTCCCGGCGATACGATGCGTGCACGCTGCCG 445

Db 301 AATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 360

Qy 446 GTCCGGCTGGCTGAGAGGAGGGGCGCGGCGCGGCGCGGCGCGGCGGCGGCGGCGG 505

Db 361 ATCGAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 420

Qy 506 GGTCCGACGACGTGCGTGGCGCGGGTGCACCGAGAGAGTAGGCATAA 556

Db 421 AATCCGAACAATACGTAACGACGCGGAATACCGAAAAAATAAATAAATAAATAA 471

RESULT 14

ABV06162/c

ID ABV06162 standard; cDNA; 303 BP.

XX AC ABV06162;

XX 13-SEP-2002 (first entry)

XX Human prostate expression marker cDNA 6153.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.

XX Homo sapiens.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US05171.

XX 17-FEB-2000; 2000US-183319P.

XX 16-MAR-2000; 2000US-189862P.

XX 25-MAY-2000; 2000US-207454P.

XX 09-JUN-2000; 2000US-211314P.

XX 18-JUL-2000; 2000US-219007P.

XX 13-DEC-2000; 2000US-255281P.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX Claim 1; Page 1018; 11750pp; English.
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX Sequence 303 BP; 81 A; 72 C; 77 G; 73 T; 0 other;
SQ
Query Match 7.4%; Score 230.2; DB 23; Length 303;
Best Local Similarity 89.7%; Pred. No. 2.5e-47;
Matches 279; Conservative 0; Mismatches 8; Indels 24; Gaps 2;
QY 1746 GTCCTTTTCAGATCCTCCGAGTCTTTCCCCCAAGTGTCTACCTATATATAATCAACCAG 1805
DB 303 GTCCTTTTCAG-TCCTCCGGTGTCTTCCCCCAAGTGTCTACCTATATATAATCAACCAG 245
QY 1806 GCTAAGACTGTATGACTCCTTTGTACTACATGTGTTCGGCATCTGTAGCTTATT 1865
DB 244 GCTAAGGACTGTATGACTCCTTTGTACTACATGTGTTCGGCATCTGTAGCTTATT 185
QY 1866 CAGATCCATGACATPAACAGGCTCGACAGAGATAGCTTGTCTATTTCTTGAGGAA 1925
DB 184 CAGATCCATGACATPAACAGGCTCGACAGAGATAGCTTGTCTATTTCTTGAGGAA 125
QY 1926 TTTCACCTTCAGGATGAGTTTATGACATTTTATTTTAAAGGACAGAGAGTTGATG 1985
DB 124 TTTCACCTTCAGGATG-AGGACAGAGAGTTGATG 88
QY 1986 CAGCGCTTCACACCATGCTGTGAAACAGGACCCCAAGGCAACATTTGGCCTGTTTAG 2045
DB 87 CAGCGCTTCACACCATGCTGTGAAACAGGACCCCAAGGCAACATTTGGCCTGTTTAG 28
QY 2046 GTACCTGGGTC 2056
DB 27 GTACCTGGGTC 17
RESULT 15
ABL07143
ID ABL07143 standard; cdNA; 2515 BP.
XX ABL07143;
XX ABL07143;
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 15911.
DE Drosophila melanogaster expressed polynucleotide; cell signalling; insecticide;
XX Drosophila; developmental biology; gene; ss.
KW pharmaceutical; gene; ss.
XX Drosophila melanogaster.
OS

XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEXE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB63040.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 15911; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2515 BP; 616 A; 715 C; 657 G; 527 T; 0 other;
Query Match 7.0%; Score 217; DB 23; Length 2515;
Best Local Similarity 47.8%; Pred. No. 1.4e-43;
Matches 1031; Conservative 0; Mismatches 1035; Indels 93; Gaps 10;
QY 639 CTGGGTGGGACATTACCCAAAGATTTTGAAGAAGCTTGTGAGAAATTAAGTTGGGAGAACT 698
DB 350 CTGGGTGGGACGTAACCAAGAGTTCCACGATGCTGCGCGGAGCTGCAGCCGGGAACT 409
QY 699 ACTTCATGATAGCTATTTGGTCTTTTGAAGCATCTGCTATTGAATGATGATC 758
DB 410 GCGCCAGGATATGCTGCTGCTGCTTTTGAAGCATCTGCTATTGAATGATGATC 469
QY 759 CAAGATGGGATGCTGGCATGATTGGAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACA 818
DB 470 CAAGATGGGATGCTGGCATGATTGGAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACA 529
QY 819 AGCTATCAAGGATGGGACATTTAAAGATTTAAAGATTTCACTTGGCTGAACTGATAGGAT 878
DB 530 AGCCATTGGCCAGCGGCGCCATCAAACTGGACGATCTCACGCCCTCCGAACTGATGGCAT 589
QY 879 TATGGATACATGTTTGTCTGTTTGAATGATGATGATGATGATGATGATGATGATGATG 938
DB 590 CTATGATGCTCTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 649
QY 939 AGTATTATGCTGCTGCTTATACATTTATACATTTATAGAGATTTATAGAGATTTATAGAG 998
DB 650 TCTATTCACTGCTCTCTACCTGACGCTCCCGCACAGATCAAGGATTAAGGCGCTGCGCT 709
QY 999 TTTTGTCTGCGGATCTTGAATCTGTGACATTTGCAAGGAAAGAAAGTAAATTAAGCTGC 1058
DB 710 CTTTCTGCACTGCGGCGGCAATCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 769
QY 1059 TGTCTTTTGAAGAGGAGGATTTTTCAGTCAATGACTTATGATTTAAATTAAGTAAAGCTG 1118
DB 770 TGTCAACGAGGAGGAGGACTTCCAGCTGTACGGAATTTCTGCTCTCTGCTGCTGCTG 829

QY 1119 GACAGATC---TTCAGTTACAGGATGCTTAAAGATGTGGAGGATGATCGCAAGAG 1175
Db 830 GGCACAGCCAGCCACTGTTTATAGCTCGTAAAGGATGTGGAGGATGAGCTCATACG--- 886
QY 1176 AGTAAAGAGTACTCGAAGTCGACAGGAGAGAAAGAGATCCAGAAAGTTGAACTAGAACA 1235
Db 887 -----CAAGTGCAAAAGCTGACTTCCACGGAGGACTCGATGG 924
QY 1236 CCAACAATGTTTACAGAGTATTCAGCAGAGTGAATTTTACTCGTGTGTTACTGACAGTGT 1295
Db 925 CGGTGTTTACCGGCTGCGTT--TCATGCGCCATCTTTTCCAGGTGATCTACCAACGTGA 982
QY 1296 TATATGCTTTTACTAAGAAAGACACAGTGTCTTCAGAGCTCAAAATTTGATGTTCA 1355
Db 983 GCAAAATGCGCAGCAATGATACCGTGCATGACAAGGTAGCATATACAAAATCCTCTTGT 1042
QY 1356 AGCAGCAGATCTTCTTTGTCATTCATAAATTCATTGTCATGTCATCCAGGCCAGAA 1415
Db 1043 GGCCTCCGAGATGCTTCCAGGAATAGGAACACTTTGGATCGGGGCACACACAGAGAA 1102
QY 1416 TGATACTACAAAAGAGATCATCCAAATTATGATGGGTTTGAACCCCTTGTGAACCCAG 1475
Db 1103 GGGGTCT-----GACGTCTCCCAATCCCATGGGTTTCTCTCGCGCATCCACGACG 1153
QY 1476 GCTACTTCCACCTACTCTTCCCTCGATATGCRAAAATAATTAAAGGAGAAATGTTGAA 1535
Db 1154 CAGCACAACCGCCGGTTCGCGTAGCATTAAGATCAGGATCGTCATCCAGTTATCA 1213
QY 1536 CTATTTTGAAGATTAATAGATAGATAAATACTGTCTGAGGTGTGTAATTTAACAA 1595
Db 1214 GTTCTAGAGAAATGATTTGCGCTTCAAAATACGCTGCAAGTCACCAAGTACAGGA 1273
QY 1596 TTTACATTTGATCTGGAATTTTCTGTGAATTTAGTGAACAGTC---ACCATGTTTCT 1652
Db 1274 TTACTANTCGCGCTGGAATCTTCTTATCGAGTACAGCAAAAGTCCGGCCAGTGCATCCT 1333
QY 1653 TTCAAGATCTCTGTTTACAAACACTTCTCTGTGTGATAACAAAAGGCTTTTGAACCTCA 1712
Db 1334 GTCCAGAGCGTGTGCAAAACCTGTTTTCAGCGCAACATGCGTATGGCGCAGGAAAGCT 1393
QY 1713 TCTCATGCAAGACATGTTGAAAGATGCACTTGGTCTTTTGTGATGATCTCCGAGTGTCT 1772
Db 1394 TCCCATGAAGCAGTTCCTCGGCCACTCGGTTTCAAGTCTTCAACTCGCCACCCGTTTGA 1453
QY 1773 TCCCCAAGTGTACTATATAATAATCACAGGCTAAGGACTGTATCGACTCTCTTTT 1832
Db 1454 T--GCCAAGATCCGGTGGCTGCCGATCCCAAGGTGCAGCAGCACTTGGAGAACTTCTTC 1511
QY 1833 ACTCACTGTGTT-----CGGCCATTTCTGTAGTCTTATTCAGATCCATGGACATAACAG 1886
Db 1512 CGTTACTGATTAAATGAACACCTTTCAGCAGTTCATTCCGATCTGCGGCTTCAATCGC 1571
QY 1887 GCTGACAGAGATAGCTTGGTTCATATCTTGAAGAAATTTGCCACTTTCAGAGATGAG 1946
Db 1572 GCCAGACAGCGGATAGCTGGCGCTCTGTAGAGAACTTCGACACCATCCAAAGTGGAT 1631
QY 1947 TTTATGACATTTTATTTAATAGGACAGAAAGTTGATGACGCGCTTTCACACCATGCTGT 2006
Db 1632 GCAGCGGATTTGGACTCTATGATGAAACCACTGGCCACGACAGGACCATGGAGGGCAAC 1691
QY 2007 TGAACAGGAACCCCAAG--GCAACATTTGCGCTTTTATAGGTACCTGGGTCTTTTACCAT 2065
Db 1692 GAGCCCATGGCCACCGCTCTGAAGCAGACAGCACCTTCTCCACATGGGTCTGTACAC 1751
QY 2066 AACCTTCGATTATGATACAGTACTCTTAAAGTGGCTTTGAATTTGAACCTCAAGTATG 2125
Db 1752 TGTTTCCGTGCGATGCTGATCTTCTGTATGCTCCGGCTTTGAGCTTGAACCTATATGCTGTG 1811
QY 2126 CACGAGTACTATTACATATTTGTTATCTCTCTGAATTCCTTTTACGATGTTTATGATCA 2185
Db 1812 CATGAGTTCCTCTACATCTACTGTTATCCATACGAGTTCCTCATTTGGCTTCCCTGCTCC 1871
QY 2186 ACATTGAGTGTGCGGATGGCTCTCAAATGGCAGAGGAAAGGATAATGGAGAGCAGCAG 2245

Db 1872 GCGCTGACCGCACCGAGAAACATCTCTCTGGCCAGGAGGAGTACGCCAGCACCAAAGT 1931
QY 2246 AAAGSCCGTAGTAG-----TAAAAAACAAG 2272
Db 1932 AAGACCCAGTCGGCGCGCAGCGGCGCTAAAAACCGCAAGCGCGCCAAAGCGAAAAAG 1991
QY 2273 AAAAAAAGAAAGTTTCGCCCATTTGAGCCGAGAGATCAAAATGAGCAAGCATATCAGAAC 2332
Db 1992 AACAAAAGACTCAAAGGCCGTACCGCCCGAGATCGTCTTCTACACGCGCTACTCAGC 2051
QY 2333 ATGTGTCTGGAATTTTAAAAACCAATGTTAGCATTTTGAATGACGGCAAGTACGTAAA 2392
Db 2052 CTCTCGCGGGAATGTATAAAGCAATGGCGCCCTGACCAAAAGACGGACGCGTGGGACTG 2111
QY 2393 CCGAAGTTTGAGCTTGATAGTGAACAAGTTTCGGTATGAACACAGCTTTTGTCTCATTCAC 2452
Db 2112 CGCTGTCCAAAGTTCGACACAGAGAGATCCGCTACAAACCGACGCTTCTGCCCCCTCGCC 2171
QY 2453 AGTGTGATGACCCCGCCCGCAGTGCATCTTACAGTTCAAGGAAATGTCTGACCTCAAT 2512
Db 2172 ACACCTCACAGTCTCTCCGCCAGTTTCGTACGAGAGTTTCAAAAACGTCCGGAGCACATG 2231
QY 2513 AAATATAGCCCTCCTCCTCAGTCTCTGAACTGTATGTGGCAGCTAGTAAAGCACTTTCAA 2572
Db 2232 A-----TCCGACATAGCGTCGAGGATCTATATACCTATGCGGCCAAAGCACTTTGAC 2282
QY 2573 CAGGCAAAAATGATATTGCAAAATATTCCTAACCCGAGCATGAGGTTTAAATAGATTTTA 2632
Db 2283 CAGGCCCAATGTTCTCGAAAGCATTCAAAACCCAGATCAGGAGATGCTGGACCTCTCG 2342
QY 2633 AAGGTTGCCAAACCAACTTTTGTGTTATGAAGTTTATTTGGCAGGAGGACACAAAAAGGAA 2692
Db 2343 CAAATCGCACGGACTAACTTTTGTGTGATGATGTCTGGCTCGCGGTCCAGAGAGGAA 2402
QY 2693 TCTAAAGTTCTCTCTGAATTTGATTTCTCTGCTCATAAATATTTTCTCTGTTGTGAAACT 2751
Db 2403 GTCAAGCGCGCAGCAGAGTTTCGACTTCTCCAAACACACAGCTACTTCCCATCATCAAGCT 2461

Search completed: November 24, 2003, 22:01:07

Job time : 790 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2003, 21:46:30 ; Search time 11138 Seconds
(without alignments)
11371.539 Million cell updates/sec

Title: US-10-001-857-42

Perfect score: 3096

Sequence: 1 ttcttcacgaactccacgg.....atttttgacaaaaaaa 3096

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.scs.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|--------|---------------|--------|----|-----------|---------------------|
| 1 | 3096 | 100.0 | 3096 | 6 | AX523454 | AX523454 Sequence |
| 2 | 2417.2 | 78.1 | 2488 | 6 | AX537452 | AX537452 Sequence |
| 3 | 1838.4 | 59.4 | 2552 | 10 | AY102701 | AY102701 Mus muscu |
| 4 | 1799.2 | 58.1 | 2543 | 10 | AF272892 | AF272892 Rattus no |
| 5 | 1766.6 | 57.1 | 1839 | 9 | AK025266 | AK025266 Homo sapi |
| 6 | 1264.8 | 40.9 | 2467 | 5 | BC053286 | BC053286 Danio rer |
| 7 | 988 | 31.9 | 997 | 9 | AK026296 | AK026296 Homo sapi |
| 8 | 885.4 | 28.6 | 1796 | 9 | AK056059 | AK056059 Homo sapi |
| 9 | 717.4 | 23.2 | 719 | 6 | AX664135 | AX664135 Sequence |
| 10 | 664.2 | 21.5 | 147876 | 2 | AC102010 | AC102010 Mus muscu |
| 11 | 551.4 | 17.8 | 181864 | 9 | AL1353743 | AL1353743 Human DNA |
| 12 | 398.4 | 12.9 | 160115 | 9 | AL161447 | AL161447 Human DNA |
| 13 | 321.4 | 10.4 | 446 | 9 | HS3434821 | HS3434821 Homo sapi |
| 14 | 299 | 9.7 | 743 | 9 | HS3434413 | HS3434413 Homo sapi |
| 15 | 236.4 | 7.6 | 642 | 9 | HS3432328 | HS3432328 Homo sapi |
| 16 | 236.4 | 7.6 | 742 | 9 | HS3431811 | HS3431811 Homo sapi |
| 17 | 233.2 | 7.5 | 640 | 9 | HS3432329 | HS3432329 Homo sapi |
| 18 | 233.2 | 7.5 | 652 | 9 | HS3432327 | HS3432327 Homo sapi |
| 19 | 217 | 7.0 | 2537 | 3 | AY060948 | AY060948 Drosophil |
| 20 | 207.6 | 6.7 | 2565 | 3 | AY089452 | AY089452 Drosophil |
| 21 | 207 | 6.7 | 320902 | 2 | AL161456 | AL161456 Homo sapi |
| 22 | 182 | 5.9 | 489 | 11 | G63197 | G63197 SHGC-140830 |
| 23 | 177.8 | 5.7 | 34935 | 9 | AL161453 | AL161453 Human DNA |
| 24 | 177.6 | 5.7 | 111554 | 2 | AC143812 | AC143812 Macaca mu |
| 25 | 155.8 | 5.0 | 113836 | 10 | AL669837 | AL669837 Mouse DNA |
| 26 | 153.2 | 4.9 | 253611 | 2 | AC111473 | AC111473 Rattus no |
| 27 | 153.2 | 4.9 | 271861 | 2 | AC108541 | AC108541 Rattus no |
| 28 | 153.2 | 4.9 | 281804 | 2 | AC134869 | AC134869 Mus muscu |
| 29 | 140.4 | 4.5 | 151 | 6 | AX523453 | AX523453 Sequence |
| 30 | 134.4 | 4.3 | 204131 | 2 | AC128898 | AC128898 Rattus no |
| 31 | 130.4 | 4.2 | 3497 | 10 | AY102702 | AY102702 Mus muscu |
| 32 | 130.4 | 4.2 | 269131 | 2 | AC135469 | AC135469 Mus muscu |
| 33 | 90.6 | 2.9 | 16016 | 2 | AC020508 | AC020508 Drosophil |
| 34 | 90.6 | 2.9 | 60135 | 2 | AC091200 | AC091200 Drosophil |
| 35 | 90.6 | 2.9 | 164443 | 3 | AC012167 | AC012167 Drosophil |
| 36 | 90.6 | 2.9 | 298827 | 3 | AE003463 | AE003463 Drosophil |
| 37 | 87.8 | 2.8 | 815 | 6 | AX523523 | AX523523 Sequence |
| 38 | 69.4 | 2.2 | 7218 | 6 | I66494 | I66494 Sequence 14 |
| 39 | 66.6 | 2.2 | 2000 | 6 | AX655393 | AX655393 Sequence |
| 40 | 66.6 | 2.2 | 110737 | 2 | AC011105 | AC011105 Homo sapi |
| 41 | 66.2 | 2.1 | 159980 | 2 | AP005743 | AP005743 Oryza sat |
| 42 | 65.6 | 2.1 | 68328 | 2 | AC137705 | AC137705 Homo sapi |
| 43 | 64.6 | 2.1 | 832 | 11 | BV026280 | BV026280 S212P6647 |
| 44 | 64.4 | 2.1 | 69373 | 2 | AC101456 | AC101456 Mus muscu |
| 45 | 64 | 2.1 | 67769 | 2 | AC068336 | AC068336 Homo sapi |

ALIGNMENTS

RESULT 1
AX523454
LOCUS AX523454
DEFINITION Sequence 42 from Patent WO02064788.
ACCESSION AX523454
VERSION AX523454.1 GI:24412350
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Macina,R.A., Recipon,H., Chen,S.Y., Sun,Y. and Liu,C.
TITLE Compositions and methods relating to lung specific genes and proteins

| | |
|----------------------------|------------------------------------------------------------------------------------------|
| JOURNAL | Patent: WO 02064788-A 42 22-AUG-2002; |
| Diadexus, Inc. (US) | |
| Location/Qualifiers | |
| source | 1. .3096 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" |
| BASE COUNT | 914 a -606 c 749 g 827 t |
| ORIGIN | |
| Query Match | 100.0%; Score 3096; DB 6; Length 3096; |
| Best Local Similarity | 100.0%; Pred. No. 0; |
| Matches 3096; Conservative | 0; Mismatches 0; Indels 0; Gaps 0; |
| Qy | 1 TTCTCAGCAACTCCGAGCGCTGTATAGAAACATAAATCCGTTCTCAGGCAGCAGTA 60 |
| Dd | 1 TTCTCAGCAACTCCGAGCGCTGTATAGAAACATAAATCCGTTCTCAGGCAGCAGTA 60 |
| Qy | 61 GCACGCTGTTGCTCTCGGAGCTTGCGTCTCGTTCGCTCGCAACCACTAAGTCTACG 120 |
| Dd | 61 GCACGCTGTTGCTCTCGGAGCTTGCGTCTCGTTCGCAACCACTAAGTCTACG 120 |
| Qy | 121 CAAACTCTCACGGTTTTCTTCGGCTTTTCGGCTTTCGGCTTCTGAAGAAATTCACAGGGCA 180 |
| Dd | 121 CAAACTCTCACGGTTTTCTTCGGCTTTTCGGCTTCTGAAGAAATTCACAGGGCA 180 |
| Qy | 181 CGCAGACGGCGGGCTCTGAGACTCCGGCTCCGCTCTTCGCCGGAACCGCCACTA 240 |
| Dd | 181 CGCAGACGGCGGGCTCTGAGACTCCGGCTCCGCTCTTCGCCGGAACCGCCACTA 240 |
| Qy | 241 CCCAGGATCCGACAGAGGGTGA AAAAGATAAACCTTCGCGTCTCGCATGCTCTTAATC 300 |
| Dd | 241 CCCAGGATCCGACAGAGGGTGA AAAAGATAAACCTTCGCGTCTCGCATGCTCTTAATC 300 |
| Qy | 301 TCGCAGAGAGAGAGCGCGCCGCTCGCGCAACGAGGCGGTGGCGAGGAGGGGGTG 360 |
| Dd | 301 TCGCAGAGAGAGAGAGCGCGCCGCTCGCGCAACGAGGCGGTGGCGAGGAGGGGGTG 360 |
| Qy | 361 TGCGCGGGAGCGGAACTCCC GGAGTAAGGAGAGGGCGGGGTTCGCGCGTCCCGG 420 |
| Dd | 361 TGCGCGGGAGCGGAACTCCC GGAGTAAGGAGAGGGCGGGGTTCGCGCGTCCCGG 420 |
| Qy | 421 GCATACGATCGCTGCAGCTCGCGGTTCGGCTGAGAGAGGGGAGGGCGGCGCG 480 |
| Dd | 421 GCATACGATCGCTGCAGCTCGCGGTTCGGCTGAGAGAGGGGAGGGCGGCGCG 480 |
| Qy | 481 GCCGAGGCGGCGTCTGTTATTTCCGTTGTCGGAACAGTGCCTGGCGCGCGGTGACCACG 540 |
| Dd | 481 GCCGAGGCGGCGTCTGTTATTTCCGTTGTCGGAACAGTGCCTGGCGCGCGGTGACCACG 540 |
| Qy | 541 GGAGAAGTAGGCATTAATGGTTATGAAGCTTCTGTAGATGATGACGATTCAAGATGGGAG 600 |
| Dd | 541 GGAGAAGTAGGCATTAATGGTTATGAAGCTTCTGTAGATGATGACGATTCAAGATGGGAG 600 |
| Qy | 601 CTCAGTATGCCAGAAAAATGGAGAAAACGAATCAAACTGGGTGGCACTTACCCAAGAT 660 |
| Dd | 601 CTCAGTATGCCAGAAAAATGGAGAAAACGAATCAAACTGGGTGGCACTTACCCAAGAT 660 |
| Qy | 661 TTGGAAGAGCTTGTGCAGAATTAAGTTGGGAGAACTACTTCAFGATAAGCTATTTGGT 720 |
| Dd | 661 TTGGAAGAGCTTGTGCAGAATTAAGTTGGGAGAACTACTTCAFGATAAGCTATTTGGT 720 |
| Qy | 721 CTTTTTGAAGCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATT 780 |
| Dd | 721 CTTTTTGAAGCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATT 780 |
| Qy | 781 GGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGSATGGCACTATT 840 |
| Dd | 781 GGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGSATGGCACTATT 840 |
| Qy | 841 AAAAAATAAGATCTCACCTTGCTGAACTGATAGGGATTAATGATACATGTTTTTTCGT 900 |
| Dd | 841 AAAAAATAAGATCTCACCTTGCTGAACTGATAGGGATTAATGATACATGTTTTTTCGT 900 |

```
QY 1981 TGATGACGCGCTTCAACACCATGCTGTGTTGAAACAGGAAACCCAAAGGCAACATTTGGCGCTG 2040
Db 1981 TGATGACGCGCTTCAACACCATGCTGTGTTGAAACAGGAAACCCAAAGGCAACATTTGGCGCTG 2040
QY 2041 TTTAGGTACCTGGGTCCTTTACCAACACCTTCGCATTTATGATACAGTACCTTCTTAAGTGG 2100
Db 2041 TTTAGGTACCTGGGTCCTTTACCAACACCTTCGCATTTATGATACAGTACCTTCTTAAGTGG 2100
QY 2101 CTTTGAATTTGGAACTCTACAGTATGACAGTACTATTACATATATTGGTATCTCTCTGA 2160
Db 2101 CTTTGAATTTGGAACTCTACAGTATGACAGTACTATTACATATATTGGTATCTCTCTGA 2160
QY 2161 ATTCTTTTACGCGATGTTGATGTCACACATTGAGTCGTGCGGATCGCTCTCAAAATGGCAGA 2220
Db 2161 ATTCTTTTACGCGATGTTGATGTCACACATTGAGTCGTGCGGATCGCTCTCAAAATGGCAGA 2220
QY 2221 GGAAGGATATGGAAGAGCGAGCAAGAGGCGGTAGTAGTAATAAATAAACAAGAAAAA 2280
Db 2221 GGAAGGATATGGAAGAGCGAGCAAGAGGCGGTAGTAGTAATAAATAAACAAGAAAAA 2280
QY 2281 GAAAGTTTCGCCATTGAGCCGAGAGATCACAATGAGCCAGCATATCAGACATGTGTGC 2340
Db 2281 GAAAGTTTCGCCATTGAGCCGAGAGATCACAATGAGCCAGCATATCAGACATGTGTGC 2340
QY 2341 TGGAAATGTTTAAACCACTGGTAGCATTTGACATGACGCGCAAGTAGCTAAACCGAAGTT 2400
Db 2341 TGGAAATGTTTAAACCACTGGTAGCATTTGACATGACGCGCAAGTAGCTAAACCGAAGTT 2400
QY 2401 TGAGCTTGATAGTGAACAAGTTTGGTATGAAACAGAGTTTGTCTCCATTCAACAGTGTGAT 2460
Db 2401 TGAGCTTGATAGTGAACAAGTTTGGTATGAAACAGAGTTTGTCTCCATTCAACAGTGTGAT 2460
QY 2461 GACCCCGCCCGCAGTCAGTACTTACAGTTTCAAGGAAATGTCTGACCTCAATAAATATAG 2520
Db 2461 GACCCCGCCCGCAGTACTTACAGTTTCAAGGAAATGTCTGACCTCAATAAATATAG 2520
QY 2521 CCTCTCTCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACCTTTCAACAGGCAAA 2580
Db 2521 CCTCTCTCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACCTTTCAACAGGCAAA 2580
QY 2581 AATGATATTGGAATAATTTCTTAACCCGAGCCATGAGTTAATAGAAATTTTAAAGTTGC 2640
Db 2581 AATGATATTGGAATAATTTCTTAACCCGAGCCATGAGTTAATAGAAATTTTAAAGTTGC 2640
QY 2641 CAAACCCAACTTTGTGTTGATGAAGTTATTGGCAGGAGGACACAAAAGGAATCTAAAGT 2700
Db 2641 CAAACCCAACTTTGTGTTGATGAAGTTATTGGCAGGAGGACACAAAAGGAATCTAAAGT 2700
QY 2701 TCCTCTGAAATTTGATTTCTCTGCTCATAAATATTTTCCTGTTGTGAAACTTGTGTTGAGA 2760
Db 2701 TCCTCTGAAATTTGATTTCTCTGCTCATAAATATTTTCCTGTTGTGAAACTTGTGTTGAGA 2760
QY 2761 GAGACTGGGAGGTGGCCATAAAGGGCAGAGTCTTCTTTTCAGACCCAACTCTTAGAGGG 2820
Db 2761 GAGACTGGGAGGTGGCCATAAAGGGCAGAGTCTTCTTTTCAGACCCAACTCTTAGAGGG 2820
QY 2821 CACATCAACAGGTCACATCAACGGAAGTGAGATGGATTTCTTGGGTAACTCACTATTA 2880
Db 2821 CACATCAACAGGTCACATCAACGGAAGTGAGATGGATTTCTTGGGTAACTCACTATTA 2880
QY 2881 TAAGGAATCTTTTAGTTGACAGCCTTATATGACATGAATGAAACTGCTGTTTAAAG 2940
Db 2881 TAAGGAATCTTTTAGTTGACAGCCTTATATGACATGAATGAAACTGCTGTTTAAAG 2940
QY 2941 TGGTTTATTATGTTCCATGAAGAACTGGTCTTATTGATGATGATGATGAAAGCTTATAT 3000
Db 2941 TGGTTTATTATGTTCCATGAAGAACTGGTCTTATTGATGATGATGATGAAAGCTTATAT 3000
QY 3001 GGTTTTATTACAGATTTAATCAAAATCATTTTATTGATGATGATGATGATGATGATGATGAT 3060
Db 3001 GGTTTTATTACAGATTTAATCAAAATCATTTTATTGATGATGATGATGATGATGATGATGAT 3060
QY 3061 TTATAAAGGTTAATAAATTTCTTGGCAAAAAA 3096
```

```
Db 3061 TTATAAAGGTTAATAAATTTCTTGGCAAAAAA 3096
RESULT 2
LOCUS AX537452 2488 bp DNA linear PAT 23-NOV-2002
DEFINITION Sequence 23 from Patent WO02070709.
ACCESSION AX537452
VERSION AX537452.1 GI:25269238
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1
AUTHORS Lal,P.G., Baughn,M.R., Yao,M.G., Walla,N.K., Elliot,V.S., Xu,Y.,
Honchell,C.D., Yue,H., Ding,L., Gietzen,K.J., Ison,C.H., Lu,D.A.,
Hafalia,A.J., Ghandi,A.R., Thangavelu,K., Sanjanwala,M.M.,
Tang,Y.T., Ramkumar,J., Griffin,J.A., Swannaker,A., Azinzai,Y.,
Sapperstein,S.K., Burford,N., Lee,E.A., Lu,Y., Tran,U.K. and
Marquis,J.P.
TITLE Molecules for disease detection and treatment
JOURNAL Patent: WO 02070709-A 23 12-SEP-2002;
Incyte Genomics, Inc. (US)
FEATURES
source
1. 2488
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/notes="Incyte ID No: 7111920CB1"
BASE COUNT 768 a 461 c 567 g 692 t
ORIGIN
Query Match 78.1%; Score 2417.2; DB 6; Length 2488;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 2470; Conservative 0; Mismatches 13; Indels 25; Gaps 2;
QY 474 GCGCGCGCGCGAGCGCGCGTTCGTTATTTCCGTGTCGGACAGTCGTCGCGCGCGGT 533
Db 1 GCGCGCGCGCGAGCGCGCGTTCGTTATTTCCGTGTCGGACAGTCGTCGCGCGCGGT 60
QY 534 GACCACGGGAGAGTAGGCATTAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 593
Db 61 GACCACGGGAGAGTAGGCATTAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 120
QY 594 ATGGGAGCTCAGTATGCGAGAAAAATGGAGAAAAACAATACAACTGGGTGGACATTAC 653
Db 121 ATGGGAGCTCAGTATGCGAGAAAAATGGAGAAAAACAATACAACTGGGTGGACATTAC 180
QY 654 CCAAGATTTGAAGAAGCTTTCGAGATTAAGTTGGGAGAACTACTTCATGATAGCT 713
Db 181 CCAAGATTTGAAGAAGCTTTCGAGATTAAGTTGGGAGAACTACTTCATGATAGCT 240
QY 714 ATTGTGCTTTTGAAGCCATGCTGCTATTGAAATGATGATCCCAAGATGATGCTGG 773
Db 241 ATTGTGCTTTTGAAGCCATGCTGCTATTGAAATGATGATCCCAAGATGATGCTGG 300
QY 774 CATGATTTGGAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 833
Db 301 CATGATTTGGAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 360
QY 834 CACTATTAATAAAGATCTCACCTTGCCTGAACTGATAGGATTTATGGATACATGTTT 893
Db 361 CACTATTAATAAAGATCTCACCTTGCCTGAACTGATAGGATTTATGGATACATGTTT 420
QY 894 TTGCTGTTTGAACGCTTGAAGCCATTTCACCTGGCACACACAGTATTTCAGTGCCT 953
Db 421 TTGCTGTTTGAACGCTTGAAGCCATTTCACCTGGCACACACAGTATTTCAGTGCCT 480
QY 954 TTACATTCATAATCCAGACTTTTATAGAAGATCTCTGATGAAGCTTTTGTCTCTGGGAAT 1013
Db 481 TTACATTCATAATCCAGACTTTTATAGAAGATCTCTGATGAAGCTTTTGTCTCTGGGAAT 540
```

QY 1014 CTTGAAATCTGTGACATTGCAAGGAAAAAGTAAATAAGCTGCTGTTTTTGAAGAGGA 1073
Db |||||
QY 541 CTTGAAATCTGTGACATTGCAAGGAAAAAGTAAATAAGCTGCTGTTTTTGAAGAGGA 600
Db |||||
QY 1074 AGATTTTCAGTCAATGACTTATGATTTAAATGGCTAACAGTGTGACAGATCTTCGAGT 1133
Db |||||
QY 601 AGATTTTCAGTCAATGACTTATGATTTAAATGGCTAACAGTGTGACAGATCTTCGAGT 660
Db |||||
QY 1134 TACAGGATGCTAAAGATGTGGAGGTGACATCCAAAGAGAGTAAAGAGTACTCCGAG 1193
Db |||||
QY 661 TACAGGATGCTAAAGATGTGGAGGTGACATCCAAAGAGAGTAAAGAGTACTCCGAG 720
Db |||||
QY 1194 TCGAACAAGGAAAGAAAGAGATCCAGAAAGTTGAATAGAACACCAACAATGTTAGCAGT 1253
Db |||||
QY 721 TCGAACAAGGAAAGAAAGAGATCCAGAAAGTTGAATAGAACACCAACAATGTTAGCAGT 780
Db |||||
QY 1254 ATTCAGCAGAGTGAATTTTACTCGTGTGTTTCTGACAGTGTGTTAGCCTTTTACAGAA 1313
Db |||||
QY 781 ATTCAGCAGAGTGAATTTTACTCGTGTGTTTCTGACAGTGTGTTAGCCTTTTACAGAA 840
Db |||||
QY 1314 AGAGACAGTCTGTGTCAGAAAGCTCAAAAATTTGATGTTTCAAGCAGCAGATCTTCTTTC 1373
Db |||||
QY 841 AGAGACAGTCTGTGTCAGAAAGCTCAAAAATTTGATGTTTCAAGCAGCAGATCTTCTTTC 900
Db |||||
QY 1374 TGCCATTCATTAATTCATTCGATCATGCGATCCAGGCCCAGAAATGATCTACAAAGGAGA 1433
Db |||||
QY 901 TGCCATTCATTAATTCATTCGATCATGCGATCCAGGCCCAGAAATGATCTACAAAGGAGA 960
Db |||||
QY 1434 TCATCCATTTATGATGGGTTTTGACCCCTTGTGACCCAGAGGCTTCCACCTACCTT 1493
Db |||||
QY 961 TCATCCATTTATGATGGGTTTTGACCCCTTGTGACCCAGAGGCTTCCACCTACCTT 1020
Db |||||
QY 1494 CCCTCGATATGCAAAAATAATTAAGGGAAGAAATGGTGAATTTTGGCAAGATTAAT 1553
Db |||||
QY 1021 CCCTCGATATGCAAAAATAATTAAGGGAAGAAATGGTGAATTTTGGCAAGATTAAT 1080
Db |||||
QY 1554 AGATAGATAAAACTGCTGTGAGGTTGTGAATTTAACTTTTACATTTGATTCCTGGA 1613
Db |||||
QY 1081 AGATAGATAAAACTGCTGTGAGGTTGTGAATTTAACTTTTACATTTGATTCCTGGA 1140
Db |||||
QY 1614 TTTTCTCTGTAATTTAGTGAACAGTCCACATGCTCTTTCAAGATCTCTGTTTACAAAC 1673
Db |||||
QY 1241 TTTTCTCTGTAATTTAGTGAACAGTCCACATGCTCTTTCAAGATCTCTGTTTACAAAC 1200
Db |||||
QY 1674 CACTTTCCTGTGTAATAAAGGCTCTTTGGAACTCATCTCATGCAAGACATGTTGAA 1733
Db |||||
QY 1734 AGATGCACTTCGGTCTTTTGTGACATCTCCGAGTCTTCCCAAGTGTCTACCTATAT 1793
Db |||||
QY 1261 AGATGCACTTCGGTCTTTTGTGACGCTCCG--GTGCTTCCCAAGTGTCTACCTATAT 1318
Db |||||
QY 1794 AATAATCACCAGGCTAAGAGCTGTATCACTCTCTTTGTTTACTCACTGTGTTTGGGCAATC 1853
Db |||||
QY 1319 AATAATCACCAGGCTAAGAGCTGTATCACTCTCTTTGTTTACTCACTGTGTTTGGGCAATC 1378
Db |||||
QY 1854 TGTAGCTTTATTCAGATCCATGGAATTAACAGGGCTCGACAGAGATGAAGCTTGGTCAAT 1913
Db |||||
QY 1379 TGTAGCTTTATTCAGATCCATGGAATTAACAGGGCTCGACAGAGATGAAGCTTGGTCAAT 1438
Db |||||
QY 1914 ATTCCTTGAGGAATTTTCCACCTTCGAGGATGATTTATGACATTTTATTTAATAGGAG 1973
Db |||||
QY 1439 ATTCCTTGAGGAATTTTCCACCTTCGAGGATG-----AGGAG 1475
Db |||||
QY 1974 AGAAGTTTGTACAGCGCTTTCACACCATGCTGTTGAAAACAGGAACCCCAAGGCAACAT 2033
Db |||||
QY 1476 AGAAGTTTGTACAGCGCTTTCACACCATGCTGTTGAAAACAGGAACCCCAAGGCAACAT 1535
Db |||||
QY 2034 TGGCCTGTTTAGGTAACCTGGGTCCTTTTACCATAACCTTCGCAATTTATGATACAGTACCTTC 2093
Db |||||
QY 1536 TGGCCTGTTTAGGTAACCTGGGTCCTTTTACCATAACCTTCGCAATTTATGATACAGTACCTTC 1595
Db |||||

QY 2094 TAACTGGCTTTGAATTCGAACTCTACAGTATGACAGTACTATTACATATATTGGTATC 2153
Db |||||
QY 1596 TAACTGGCTTTGAATTCGAACTCTACAGTATGACAGTACTATTACATATATTGGTATC 1655
Db |||||
QY 2154 TCTCTGAATTCCTTTACGATGTTGATGTCAAATTTAGTCTGTCGCCGATGGCTCTCAAA 2213
Db |||||
QY 1656 TCTCTGAATTCCTTTACGATGTTGATGTCAAATTTAGTCTGTCGCCGATGGCTCTCAAA 1715
Db |||||
QY 2214 TGGCAGAGGAAGGATTAATGGAAGCAGCAGCAAGAGCCGTAGTAGTAAATAAAACAAAGA 2273
Db |||||
QY 1716 TGGCAGAGGAAGGATTAATGGAAGCAGCAGCAAGAGCCGTAGTAGTAAATAAAACAAAGA 1775
Db |||||
QY 2274 AAAAAAAGAAAGTTTCGCCCATTTGAGCCGAGAGATCACAATGAGCCAAACATATCAGAAAC 2333
Db |||||
QY 1776 AAAAAAAGAAAGTTTCGCCCATTTGAGCCGAGAGATCACAATGAGCCAAACATATCAGAAAC 1835
Db |||||
QY 2334 TGTGTCTGGAATGTTTAAAAACCATGTTAGCATTTGACATGGAACGGGAAAGTACGTAAAC 2393
Db |||||
QY 1836 TGTGTCTGGAATGTTTAAAAACCATGTTAGCATTTGACATGGAACGGGAAAGTACGTAAAC 1895
Db |||||
QY 2394 CGAAGTTTGAAGTGTAGTGAACAAAGTTTCGGTATGAACACAGGTTTCTCCATTCAACA 2453
Db |||||
QY 1896 CGAAGTTTGAAGTGTAGTGAACAAAGTTTCGGTATGAACACAGGTTTCTCCATTCAACA 1955
Db |||||
QY 2454 GTGTGATGACCCGCCGCCAGTGCACCTTTACAGTTTCAAGAAATGTTGACCTCAATA 2513
Db |||||
QY 1956 GTGTGATGACCCGCCGCCAGTGCACCTTTACAGTTTCAAGAAATGTTGACCTCAATA 2015
Db |||||
QY 2514 AATATAGCCCTCCCTCCAGTCTCTGAACTGTATGTTGGAGCTAGTAAAGCACTTTCAAC 2573
Db |||||
QY 2016 AATATAGCCCTCCCTCCAGTCTCTGAACTGTATGTTGGAGCTAGTAAAGCACTTTCAAC 2075
Db |||||
QY 2574 AGGCAAAATGATATTTGGAAATATTCCTAACCCGACCATGAGGTTAATAGAAATTTTAA 2633
Db |||||
QY 2076 AGGCAAAATGATATTTGGAAATATTCCTAACCCGACCATGAGGTTAATAGAAATTTTAA 2135
Db |||||
QY 2634 AGGTTGCCAAACCCAACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAAT 2693
Db |||||
QY 2136 AGGTTGCCAAACCCAACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAAT 2195
Db |||||
QY 2694 CTAAAGTTCCCTCCGAAATTTGATTTCTCTGCTCATTAATATTTTCTGTTGAACTTG 2753
Db |||||
QY 2196 CTAAAGTTCCCTCCGAAATTTGATTTCTCTGCTCATAAATATTTTCTGTTGAACTTG 2255
Db |||||
QY 2754 TTTGAGAGAGACTGGGAGGTGGCCATAAAGGGGAGAGTCTTCTTTCAGACCCCAACTCT 2813
Db |||||
QY 2356 TTTGAGAGAGACTGGGAGGTGGCCATAAAGGGGAGAGTCTTCTTTCAGACCCCAACTCT 2315
Db |||||
QY 2814 TAGAGGGCAGATCAGCAGGCTCCACATCAGCGGAAGTGAAGTGGATTCTTGGGTAAACA 2873
Db |||||
QY 2316 TAGAGGGCAGATCAGCAGGCTCCACATCAGCGGAAGTGAAGTGGATTCTTGGGTAAACA 2375
Db |||||
QY 2874 CTCAATTAAAGGAATCTTTTAGTTTGAACGCTTTATATGACATGAATGAAACTGCTGT 2933
Db |||||
QY 2376 CTCAATTAAAGGAATCTTTTAGTTTGAACGCTTTATATGACATGAATGAAACTGCTGT 2435
Db |||||
QY 2934 TTTAAAGTGGTTTATTTATGTTTCCATGGAAGAACTGGTCTTTATTGAAT 2981
Db |||||
QY 2436 TTTAAAGTGGTTTATTTATGTTTCCATGTAAGACACTGGGTTCCATTAT 2483
Db |||||

RESULT 3

AY102701

LOCUS

DEFINITION

AY102701

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

| REFERENCE | 1 (bases 1 to 2552) | Query Match | 59.4%; | Score 1838.4; | DB 10; | Length 2552; |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|--------------|--------------------------------------|------------|--------------|
| AUTHORS | Wenzlau,J.M. and Weiser-Evans,M.C.M. | Best Local Similarity | 84.4%; | Pred. No. 0; | | |
| TITLE | Mouse Embryonic Growth Associated Protein (EGAP) | Matches 2160; | Conservative | 0; <th>Mismatches</th> <td>341;</td> | Mismatches | 341; |
| JOURNAL | Unpublished | | | | Indels | 57; |
| REFERENCE | 2 (bases 1 to 2552) | | | | | Gaps 6; |
| AUTHORS | Wenzlau,J.M. and Weiser-Evans,M.C.M. | | | | | |
| TITLE | Direct Submission | | | | | |
| JOURNAL | Submitted (13-MAY-2002) Pediatrics, University of Colorado HSC, | | | | | |
| FEATURES | 4200 East Ninth Ave, Denver, CO 80262, USA | | | | | |
| source | Location/Qualifiers | | | | | |
| | 1. .2552 | | | | | |
| | /organism="Mus musculus" | | | | | |
| | /mol_type="mRNA" | | | | | |
| | /db_xref="taxon:10090" | | | | | |
| | /chromosome="11" | | | | | |
| | /dev_stage="embryo" | | | | | |
| | 55. .2232 | | | | | |
| | /codon_start=1 | | | | | |
| | /product="embryonic growth-associated protein EGAP" | | | | | |
| | /protein_id="AAM52342.1" | | | | | |
| | /db_xref="GI:21539896" | | | | | |
| | /translation="MVKATVDDDASGWEIVGPEKMEKSSWVDITQDPEDACREKLK LGLLHDLKFLFEAFSAIEMMDPKMDAGINQVNRKVLNFEQATKDGTIKIKDLSL PELIGIMDTCFCCLIYTWEGHSLAQTFTCLYTHNDFIEDPAKFAALGILKIDIA RKVNKAAVAPEEDFQSMVTGFKMANSVTDLRVTGMLKQVEDDLQBRVKSTRSQGEE RDEVELEHOQCCLAAASRVKFTFVLLTLLIAFTKETSVAEAKLKVQADLLLSAIH TSLHGHAQNGTQTDGHPILMNGPEPLNORLLPPTPRYAKIIKKEEMVYESRLID RIKTCEVSNVNLNMLHCLIDDFCSBQSPCLSRSLQTLTFLVDNRKHFVTHLMQDMY KDALSRSVPVLSKCLLYNNQAKDCIDSVFTHCVRPFCSLVQKHGHRARQDKL GHLEEFATQDQAEKVFDAFLWMLLKQBPOROHACLGTWVLYNLRITWIOVLLSGF ELELYSHKEYIYLYWVLSFEFLAWLMSLSDRAGSOWAEEIIMEEQKGSSKKTKK KYRPLSREITMSQAYQNMCAKMFKNVAFDMQGVKPKPFELDSQVRYTEHRFAFPN KVMTPPVKIQCFKEMSLDGFPPQSPSLYVAASKHFOOTKMLLENINPNPDREVS ILKVFNFVVMKLKAGHKKESKVPFEFDFSVHKYFPVVKLV" | | | | | |
| BASE COUNT | 740 a 542 c 619 g 651 t | | | | | |
| ORIGIN | | | | | | |
| | Query Match | | | | | |
| | Best Local Similarity | | | | | |
| | Matches 2160; | | | | | |
| | Conservative | | | | | |
| | 0; | | | | | |
| | Mismatches | | | | | |
| | 341; | | | | | |
| | Indels | | | | | |
| | 57; | | | | | |
| | Gaps | | | | | |
| | 6; | | | | | |
| Qy | 541 GGAAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGAGTCAGGATGGAG 600 | | | | | |
| Db | | | | | | |
| | 40 GCGCGAGCGCATCATGGTATGAAGACTACTGTAGACACGATGCTTCGGATGGAG 99 | | | | | |
| Qy | 601 CTCAGTATGCCAGAAAAATGGAGAAAGCAATACAAACTGGGTGGACATTACCCAGAT 660 | | | | | |
| Db | | | | | | |
| | 100 CTCGGGTCCGGAAAAAATGCAAAAAGTAGCACAGCTGGGTGGACATAACCCAGAT 159 | | | | | |
| Qy | 661 TTTGAAGAAGCTGTCGAGATTAAAGTCTGGGAGCACTACTTCATGATAGCTATTGGT 720 | | | | | |
| Db | | | | | | |
| | 160 TTTGAAGATGCTTGTGAGAGCTGAAGTGGGAGACTGCTTCATGATAAGCTGTTTGGT 219 | | | | | |
| Qy | 721 CTTTTTGAAGCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATT 780 | | | | | |
| Db | | | | | | |
| | 220 CTTTTTGAAGCCATGCTGCTATTGAAATGATGGATCCTTAAGATGGATGCCGTATGATC 279 | | | | | |
| Qy | 781 GGAAACCAAGTTAATCGAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATT 840 | | | | | |
| Db | | | | | | |
| | 280 GGAAACCAAGTGAATAGAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATT 339 | | | | | |
| Qy | 841 AAAATTAAAGATCTCACCTTGCCCTGAACTCATAGGATTTATGATCATGCTTTTCTGCTGT 900 | | | | | |
| Db | | | | | | |
| | 340 AAAATTAAAGACCTCACGCTTGCCCTGAACTCATAGGATTAATGACACCTGTTTCTGCTGT 399 | | | | | |
| Qy | 901 TTGATAACGTTGTTAGAAAGCCATTCCTCTGGCACAGACAGTATTATACGTGCCTTTACATT 960 | | | | | |
| Db | | | | | | |
| | 400 TTGATCACATGGCTCGAAGSCCATTCCTTGGCACACAGTGTATTACGTGCCTTTACATT 459 | | | | | |
| Qy | 961 CATATCCACACTTATAGAAGATCTCGTATGAAGCTTTTTCGCTCTGGGAATCTTTGAAA 1020 | | | | | |
| Db | | | | | | |
| | 460 CATATCCCAACTTCATAGAAGATCTCGCATGAAGACTTTTTCGCTCTGGGAATCTTTGAG 519 | | | | | |

QY 910 TGGTTAGAGGCCATTCACTGGCAGACAGAGTATTACGTGCTTTACATTCAATAATCCA 969
DB |||||
418 TGGCTCGAAGGCCATTCTCTGGCAGACAGCGTATTACATGCTTTACATTCAATAATCCA 477
QY 970 GACTTTATAGAAAGATCTGTCTATGAAAGCTTTTGTCTGTGGAATCTTGAAATCTGTGAC 1029
DB |||||
478 GACTTTCATAGAAAGACCTCGCCATGAAAGCTTTTGTCTGTGGAATCTTGAAGATCTGTGAC 537
QY 1030 ATTCCAGGGGAAAAGATAAATAAGCTGCTCTTTTGAAGAGGAGATTTTCAGTCAATG 1089
DB |||||
538 ATTCCAGGGGAAAAGATCAAGAGCTGCTCTTTTGAAGAGGAGATTTTCAGTCAATG 597
QY 1090 ACTTATGGATTTAAAATGGCTTAAACAGTGTGACAGATCTTCCAGTTACAGGCATGCTMAAA 1149
DB |||||
598 ACATATGGATTTAAAATGGCCAAACGGTGTGACAGATCTTCCAGTTACAGGCATGCTCAA 657
QY 1150 GATGTGGAGGATGACATGCAAAAGAGATTAAGAGTACTCGAAGTGGCAAGAGAGAA 1209
DB |||||
658 GATGTGGAGGACGACATGCAAAAGAGAGTAAGAGTACTCGAAGTGGCAAGAGAGAGAG 717
QY 1210 AGAGATCCAGAAAGTTGAACTAGAACACCAACAATGTTTTAGCAGTATTCCAGCAGAGTGAAA 1269
DB |||||
718 AGAGACCCAGAGTTGAATCTAGAACACACAGCAGTGTCTGGCAGCGTTCCAGCAGAGTGAAG 777
QY 1270 TTTACTCGTGTGTTACTGACAGTGTCTATAGCCTTTTACTAAAGAAAGACCAAGTGTCTGTT 1329
DB |||||
778 TTCACCCGAGTACTGCTGACAGTGTCTATAGCCTTCACTAAGAAAGAACCAAGTGTCTGTT 837
QY 1330 GCAGAGCTCAAAAATTTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCATTAATCA 1389
DB |||||
838 GCAGAGGCTCAGAAATCTGATGGTCCAGGCGCAGACCTTCTTCTGCGCAATTCACACCTCA 897
QY 1390 TTGCATCATGCGATCCAGGCCAGAAATGATACTACAAAAGGAGATCATCCCAATTTATGATG 1449
DB |||||
898 TTGCACACGCGATCCAGGCTCAGAAATGCGACTACCAAGAGAGACCATCCCAATTTATGATG 957
QY 1450 GGTTTTGAACCCCTTGTGAACACAGAGGCTACTTCCACCTACCTTCCCTCGATATGCAAAA 1509
DB |||||
958 GGTTTTGAGCCCTTGTAAACAGAGATTAATCTCCACCCACTTCCCTCGCTATGCAAAA 1017
QY 1510 ATAAATTAAGGGAGAAATGGTGAATTTAACAATTTAATGATGATACCTCTGCTGGAATTT 1569
DB |||||
1018 ATAAATTAAGAGAGAAATGGTCAACTATTTCTCAAGATTAATAGACAGAGATTAATAACT 1077
QY 1570 GTCTGTGAGGTGTGAAATTAACAATTTAACAATTTAATGATGATACCTCTGCTGGAATTT 1629
DB |||||
1078 GTCTGTGAGGTGTGAAATTAACAATTTAACAATTTAATGATGATACCTCTGCTGGAATTT 1137
QY 1630 AGTGAACAGTCAATGTGTTCTTTCAAGATCTCTGTTPACAAACCACTTTCTCTGTGAGAT 1689
DB |||||
1138 AGTGAACAGTCAATGTGTTCTTTCAAGATCTCTGTTPACAAACCACTTTCTCTGTGAGAT 1197
QY 1690 AACAAAAGGCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACTTCGGTCT 1749
DB |||||
1198 AACAAAAGGCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCTTCCCGGTCC 1257
QY 1750 TTTGTGAGTCCCTCCGAGTGTCTTTCCCGCAAGTGTCTATATAATAATCAACAGGCTA 1809
DB |||||
1258 TTTGTGAG--TCTCTCCGTTGCTCTCCCGCAAGTGTCTGCTGTATATAATCAACAGGCTA 1315
QY 1810 AGGACTGTATGACTCTCTTTGTTACTCACTGTGTGGGCCATCTGTAGTCTTATTCAGA 1869
DB |||||
1316 AGGACTGTATGACTCTCTTTGTTACTCACTGTGTGGGCCATCTTTGTAGTCTTGTTCAGA 1375
QY 1870 TCCATGACATAACAGGGCTCGACAGAGAGATAGCTTGGTCAATTTCTTGAGAAATTTG 1929
DB |||||
1376 TCCATGAGCAATAACAGGGCTCGGACAGAGAGACAAACTTGGTCAATTTCTTGAAGAGTTG 1435
QY 1930 CCACCTTGCAGGATGAGTTATGACATTTATTTAATAGGCAGAGAGGTTGATGCAGC 1989
DB |||||
1436 CCACCTTGCAGATG-----AGGCAGAGAGGTCGATGCAGC 1472

QY 1990 GCTTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATTTGGCTGTTTAGGTAC 2049
DB |||||
1473 GCTCCAACATATGTTGTTGAAGCAGGAGCCTCAGCAGACAGATCTGGCTGTTTAGGAAC 1532
QY 2050 CTGGGTCTTTTACCAATAACTTTCGCAATATGATACAGTACCTTCTTAAGTGGCTTTGAAT 2109
DB |||||
1533 TTGGGTTCTCTACACAGCCTGCGAATATGATCCAGTACCTGCTCAGCGGCTTTGATCT 1592
QY 2110 GGAATCTACAGTATGACAGGATCTATTAATATATTTGGTATCTCTCTGAATTTCTTTA 2169
DB |||||
1593 GGAGCTCTACAGCATGACAGATCTACTACATCTACTGCTGCTCTCTGAGTCTCTGTA 1652
QY 2170 CGCATGCTGATGTCACATGAGTCTGCGCATGCGCTCTCAATGGCAGAGGAAAGAT 2229
DB |||||
1653 CGCATGCTGATGTCACCTGAGCCGTGCTGACGGCTCTCAGATGGCAGAGAGGAT 1712
QY 2230 AATGGAAGAGCAGAGAAAGCCGTAGTAGTAAAAAAAACAAGAAAAAAGAAAGTTGG 2289
DB |||||
1713 AATGGAAGAAACAGCAGAAAGGCCGAGCAGCAAGAAAAAACAAGAAAAAAGAAAGTTGG 1772
QY 2290 CCCATTGAGCCGAGAGATCAATAGCCAGCATATCAGAACATGCTGCTGGAATGTT 2349
DB |||||
1773 TCCATTGAGCCGAGAGATCAATAGCCAGCCCTATCAGAACATGCTGCTGGGATGTT 1832
QY 2350 TAAACCATGCTAGCATTTGACATGGACGCAAAAGTACGTAAACCGAAAGTTTGACTTGA 2409
DB |||||
1833 CAAACCATGCTGCAATTTGACATGGATGGCAAGTGGCAAGCCCAATTTGACTTGA 1892
QY 2410 TAGTGAACAAGTTCCGTTATGAAACAAGTTTTGCTCCATTTCAACAGTGTGATGACCCGCC 2469
DB |||||
1893 CAGTGAGCAGTCCGCTATGAGCAGAGTTTGGTCCGTTCAACAGCGTGTGACGCCACC 1952
QY 2470 GCCAGTCACTACTTACAGTTCAAGGAATCTGTGACCTCAATAAATATAGCCCTCTCC 2529
DB |||||
1953 ACCAGTCACTACTTACAGTTCAAGGAATCTGTGACCTCAATAAATATAGCCCTCTCC 2012
QY 2530 TCAGTCTCTCTGAACTGTATGTGGCAGCTAGTAAAGCACTTTCAACAGGCAAAATGATATT 2589
DB |||||
2013 TCAGCTCCAGAGCTGTATGTGGCAGCTAGTAAAGCACTTTCAACAGGCAAAAGATGATCT 2072
QY 2590 GGAATAATTTCTTAACCCGCAATAGAGTTAAATAGAAATTTAAAGTTGCCAACCCAA 2649
DB |||||
2073 TGAGAGCATCCCAATGACAGCCGGAGTTCAGCAGGATCTTAAAGTTGCCAAGCCATA 2132
QY 2650 CTTTGTGTTATGAGTTATTTGGCAGGAGGACACAAAGAGAAATCTAAAGTTCTCTCTGA 2709
DB |||||
2133 CTTTGTGTTATGAGGCTTTTAGCAGGAGGACACAAAGAAAGTCAAAAGTTCTCTCTGA 2192
QY 2710 ATTTGATTTCTCTCTCATATAATTTTCTCTGTTGTGAAACTTGTTCAGAGAGACTGG 2769
DB |||||
2193 ATTTGATTTCTCTCTTACAAATATCTTCTGTTGTGAAACTGTTTGAAGAGACTGCA 2252
QY 2770 GAGTGGCCATAAAGGGGACAGAGTCTTTCTTTACAGCCCACTCTTAGAGGGACATACCC 2829
DB |||||
2253 AAGATGACCAT-----GTCTACTCTGGGGTACACCA 2283
QY 2830 AGGCTCCACATCAGGGAAGTGAGATGGAATTTCTTGGGTAACTCACTATTAAGGA--- 2886
DB |||||
2884 GTGFGCCACACACAGGTGACATAAAGTGGCTCTCTTTGAATGACAGCTTTGTTACAGAGAC 2343
QY 2887 -ATACTTTTATGTTGACAGCTTTATATGACATGAATGAATACTGCTGTTTAAAGTGGTT 2945
DB |||||
2344 TTCAGTATGATCAACACCTTACCTGGAGAGACG-AAAAGTCTGTTTAAAGTGGTT 2402
QY 2946 TATTATGTTCCATGGAAAGAACTGGTCTTATGATGCAATGATGAACAGTTTATATGGTTT 3005
DB |||||
2403 TGTATATCTCCATGGGTGTGACGGG----CTGGAATGCAITGGTGAACGTTTACATGGTTT 2458
QY 3006 TATTACAGATTTAATCAAAATCAATTTTATGAAATGATGAGTGAAGAAATAGTGTATTATA 3065
DB |||||
2459 TATTACAGATCTTCAATGTAAGAACTTTTAAATGAATGATTTGTG-AAAATAGTGTTTGTA 2517
QY 3066 AAGGTTAATAAATTTCTTGACAAAA 3091


```
Db 1141 AAAAAGCTGTCTGAGGTTGTGAATTTTAAACAAATTTACATTTGTATCTCTGGATTTTCTG 1200
Qy 1623 TGAATTTAGTGAACAGTCCACATGCTGTTCTTCAAGATCTCTGTACAAACCACTTTCCCT 1682
Db 1201 TGAATTTAGTGAACAGTCCACATGCTGTTCTTCAAGATCTCTGTACAAACCACTTTCCCT 1260
Qy 1683 GGTGGATAACAAAGAGTCTTTTGAACATCTCTCATGCAAGACATGTTGAAAGATGCACT 1742
Db 1261 GGTGGATAACAAAGAGTCTTTTGAACATCTCTCATGCAAGACATGTTGAAAGATGCACT 1320
Qy 1743 TCGGTCTTTTGTAGATCTCCGAGTGTCTTCCGCCAAGTGTCTACCTATATATAATCAAC 1802
Db 1321 TCGGTCTTTTGTAG--TCCTCCGGTGTCTTCCGCCAAGTGTCTACCTATATATAATCAAC 1378
Qy 1803 CAGGCTAAGGACTGTATCGACTCTTTGTACTCACTGTGTTCGGCCATCTGTAGTCTT 1862
Db 1379 CAGGCTAAGGACTGTATCGACTCTTTGTACTCACTGTGTTCGGCCATCTGTAGTCTT 1438
Qy 1863 ATTCAAGTCCATGACATACAGGCTCGACAGAGATAGCTTGTTCATTTCTTTGAG 1922
Db 1439 ATTCAAGTCCATGACATACAGGCTCGACAGAGATAGCTTGTTCATTTCTTTGAG 1498
Qy 1923 GAATTTCCACCTTCGAGGATGCTTTATGACATTTTATATAGCAGAGAGGTTG 1982
Db 1499 GAATTTCCACCTTCGAGGATG-----AGCAGAGAGGTTG 1535
Qy 1983 ATGCAAGGCTTTCACCATGCTGTGTGAACAGGACCCCAAGGCAACATTTGGCCCTGTT 2042
Db 1536 ATGCAAGGCTTTCACCATGCTGTGTGAACAGGACCCCAAGGCAACATTTGGCCCTGTT 1595
Qy 2043 TAGGTACTCGGTGCTTTACCAATACCTTCGATATGATACAGTACCTTCTAAGTGCT 2102
Db 1596 TAGGTACTCGGTGCTTTACCAATACCTTCGATATGATACAGTACCTTCTAAGTGCT 1655
Qy 2103 TTGAATTTGAACTCTACAGTATGACGAG--TACTATTACATATATTGGTATCTCTCGAA 2161
Db 1656 TTGAATTTGAACTCTACAGTATGACGAGTACTATTACATATATTGGTATCTCTCGAA 1715
Qy 2162 TTCCTTTACGATGTTGTATGATCAACATTTAGTGTGCTGCGGATGCTCTCAATGGCAGAG 2221
Db 1716 TTCCTTTACGATGTTGTATGATCAACATTTAGTGTGCTGCGGATGCTCTCAATGGCAGAG 1775
Qy 2222 GAAAGATAATGGAAGAGCAGAGAGAGGCGGTAGTAGTAAAGGAGGAGGAGGAGGAGGAG 2281
Db 1776 GAAAGATAATGGAAGAGCAGAGAGAGGCGGTAGTAGTAAAGGAGGAGGAGGAGGAGGAG 1835
Qy 2282 AAA 2284
Db 1836 AAA 1838

RESULT 6
BC053286
LOCUS BC053286 2467 bp mRNA linear VRT 09-JUN-2003
DEFINITION Danio rerio cdna clone MGC:64157 IMAGE:6797204, complete cds.
ACCESSION BC053286
VERSION BC053286.1 GI:31418975
KEYWORDS MGC.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 2467)
Klausner,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Strausberg,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
```

```
TITLE
JOURNAL
MEDLINE
PUBMED
2 (bases 1 to 2467)
Strausberg,R.
Direct Submission
Submitted (02-JUN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Leonard I. Zon, M.D.
cdna Library Preparation: Invitrogen Corp.
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 117 Row: n Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Genomescan gene prediction, Similarity but not identity
to protein.

Location/Qualifiers
1. 2467
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="MGC:64157 IMAGE:6797204"
/tissue_type="Kidney, zebrafish"
/clone_lib="NCI CGAP_Zklid"
/lab_host="DH10B"
/notes="vector: pCMV-SPORT6.1"
58..2232
/codon_start=1
/product="Unknown (protein for MGC:64157)"
/protein_id="AAH53286.1"
/db_xref="GI:31418976"
/translation="MWMKSSVEEEEGCGGLGPEKMRNNANWVDVTFKGCACKELKL
GELHDKLGLFPEAMSAENWDFNDAGMIGNQVRKVLNFDQAVKDEALVKDLSIP
ELIGIMDTCCCLITWLEHSLAQVFTCYLHVHNPDLIQDPALKAFLGILKIDIAIR
EKVNAKAFEEEDFQAMTYGFKMANNVTDLRVTGMLKXDEQLQKVKSTRSQGEQR
DPVELDHQOCLALFKFKFTRLLSALISFTKKTSAVSAEQKLMQOALLPAVHA
TIQVIGSQNDTTGGDHPIMWGPELVNORLLPTTPRYAKIKREEMVNVFSLIER
IKSCVEVINITNLHSILDPFCESEQSCVLSLLQTTTLINNKVFGTHLQMDMIK
DALRCVSPVLSSKSLNNHQAQIDISFVTHCTRFPCSLYQHGNRQRQDKLG
HILFEFATQDEAKVDAAHLGLMLKLEPQRLHCLGTWILVHNLRIHTQYLSSGP
LEYSHEYYIYWIYLSFELYAWLMTSLRSDSSQMAEERILBEQLKVRSSKKSKK
KAPLSKEITMSQAYQNMCAQYKMTALMDRKRKPKQFELDSQVRYBHRPAPFNS
VTPPPVHYIQKEMSDLKYNPPPSADLYMAASKHFQQAOKLLENVTFDABVNI
LKVAQNVIMVLLAGGCHKETKALPEPDSAKHYPIVKIL"
BASE COUNT 720 a 574 c 600 g 573 t
ORIGIN

Query Match
40.9%; Score 1264.8; DB 5; Length 2467;
```

| | | | |
|----------------------------------------------------------------------------------------------------------------------|------|---------------------------------------------------------------|------|
| Best Local Similarity 74.3%; Pred. No. 8.3e-276; Matches 1648; Conservative 0; Mismatches 542; Indels 28; Gaps 3; | | | |
| Qy | 543 | AGAAAGTGGATTAATGGTTATGAAGCTTCTGTAGATGATGACGATTCAGGATGGAGCT | 602 |
| Db | 45 | AAAGAGCTCTCCATGGTGATGAAGTCAATCGGTGAGGAGGAGGAGATGGGTCT | 104 |
| Qy | 603 | CAGTATGCCAGAAAAAATGGAAGAAAGCAATACAACTGGGTGGACATTACCCAGATTT | 662 |
| Db | 105 | TGGGATCCAGAGAAGAT--GAGAAACAATGTCCTGGGTGGACGTCACCCAGAGTT | 161 |
| Qy | 663 | TGAAGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCAATGATAGCTATTGGTCT | 722 |
| Db | 162 | TAAAGGAGCTTGTAAAGAACTCAAACTTGGCGAGCTGCTCCATGCAAGCTATTCCGGTCT | 221 |
| Qy | 723 | TTTTGAAGCCATGCTGCTATTGAATGATGATCCCAAGATGGATGCTGGCATATTGG | 782 |
| Db | 222 | GTTCGAAGCCATGTCGCAATGATGATGATCGAAGATGGATGCTGNAATCGG | 281 |
| Qy | 783 | AAACCAAGTTAATCGAAAAAGTTCTCAATTTGAAACAAGCTATCAAGGATGGCACTATTAA | 842 |
| Db | 282 | AAATCAAGTCAACCGTAAAGTCTCAACTTTGACCAAGCTGTTAAGGATGAAGCCATACG | 341 |
| Qy | 843 | AAATTAAGATCTCACCTTGCCTGAACTGATAGGATATGATACATGTTTTCGTGTT | 902 |
| Db | 342 | GGTGAAGATCTAAGTATTCCTGAGCTCATCGGATCATGGACATGTTTCTGCTGCT | 401 |
| Qy | 903 | GATAACGTGGTTAGAAAGCCATTCACTGGCAAGACAGTATTATTCGTGCTTTTACATCA | 962 |
| Db | 402 | GATCACGTGGCTGGAGGGCACTCTCGCTCAGACCGCTTTACATGCTGTAGCTTCA | 461 |
| Qy | 963 | TAATCCAGCTTTAAGAAAGTCTGCTATGAAGGCTTTGCTCGGAACTCTGAAAT | 1022 |
| Db | 462 | TAATCCGATCTGATTTCAAGACCTTCGCTCAAAGCCTTCGCTCGGCACTCCTGAAAT | 521 |
| Qy | 1023 | CTGTACACATTCGAAGGAAAAAGTAAATAAGCTGCTGTTTTGAAGAGGAAATTTTCA | 1082 |
| Db | 522 | CTGTACATCGCCGTGAAGAAAGTCAACAACTGCTGCTGCTGCTGCTGCTGCTGCT | 581 |
| Qy | 1083 | GTCATGACTTATGGAATTAATAATGGCTAACAGTGTGACAGATCTTCGAGTTTACAGCAT | 1142 |
| Db | 582 | GGCCATGACTTACGCTTTAAATGGCCAAACAACTAATGATCTGAGAGTCAAGGAT | 641 |
| Qy | 1143 | GCTAAAAGATGTGGAGGATGACATGCAAGAAAGATTAAGAGTCTGAAAGTCGAAGG | 1202 |
| Db | 642 | GCTGAAGATGTGAGGATGAACCTTCAGAGGAAAGTAAAGATACCCCGAGTCGACAGG | 701 |
| Qy | 1203 | AGAAAGAAAGATCCAGAAAGTTGAATAGAACACCAACAATGTTTAGCAGTATTCAGCAG | 1262 |
| Db | 702 | TGAACAGCGAGATCTGAGTGGAACTGGATCATCAGCAGTCTTGGCTCTGTTGAGTCG | 761 |
| Qy | 1263 | AGTGAATTTACTCGTGTGTTTACTGACAGTGTCTTATPAGCCTTTAATAAGAAAGACCCAG | 1322 |
| Db | 762 | GGTTAAATTTCACTGCTGCTGCTCAGCGCGCTAATCTCTTTCAGCAAAAAGAGACGAG | 821 |
| Qy | 1323 | TGCTGTGACAGAGCTCAAAAATGATGTTCAAGCAGCAGATCTCTTTCGCTGCAATCA | 1382 |
| Db | 822 | TGCTGTGACGAGACAGAACTGATGATCAAGTCTGATTTACTGCTCTGAGTGTCA | 881 |
| Qy | 1383 | TAAITCAITTGATCATGGCATCCAGGCCAGAAATGATACATAAAGAGAGATCATCCAAAT | 1442 |
| Db | 882 | TGCTACCATACAGTACGGCATACAGTCGAGAACGACACTCAAAAGGAGATCACCCGAT | 941 |
| Qy | 1443 | TATGATGGGTTTTGAACCCCTTGTGAAACAGAGGCTACTTCCACCTTACCTTCCCTCGATA | 1502 |
| Db | 942 | CATGATGGGTTTTGAGCCTCTGGTTAAACAGCGCCTGCTCCACCTCACTGCTCCCGGATA | 1001 |
| Qy | 1503 | TGCAAAATTAATAAAGGAAAGAAATGGTGAATTTTGAAGTAAATGATAGAT | 1562 |
| Db | 1002 | TCCCAAGATCATACAGCGAGAGAGATGGTCAACTACTTTCAGCAAACTCAITGAGCGCAT | 1061 |
| Qy | 1563 | AAAAAAGTGTGTGAGGTTGTGAATTTTAAACAAATTTACATTTGATCTCTGGATTTTCTG | 1622 |

| | | | |
|----|------|----------------------------------------------------------------|------|
| Dh | 1062 | TAAAGAGCGTGTGAGGTGATCAACATCAACCACTGCACAGATCTCTGGATTTCTTCTG | 1121 |
| Qy | 1623 | TGAATTTAGTGAACAGATCACCATGTGTTCTTTTCAAGATCTCTGTTTACAAACCACTTTCT | 1682 |
| Db | 1122 | TGAATTCAGTCAGCAGTCTCCATGTGTCTGTCTAGATCTCTACTGCAGACCCACCTTCT | 1181 |
| Qy | 1683 | GGTGATAAACAAAAGGTCTTTGGAATCATCTCTCATGCAACATGTTGGAAGATGCAT | 1742 |
| Db | 1182 | GATCGAACAAAGAGGTGTTGGGACTCATCTGATGAGGATATGATTAAAGATGCGCT | 1241 |
| Qy | 1743 | TCGGTCTTTTCTCAGATCCTCCGAGTGTCTTCCCAAGTGTACTATATAATATAC | 1802 |
| Db | 1242 | GCGCTGTTTTGTGAG--TCCACCTGTTCTTCTCCAAATGCAGTTTGAATAAACCA | 1299 |
| Qy | 1803 | CAGGCTAAGGACTGTATCGACTCTCTTTTACTCATCTGTTTCGCCCACTTCTGTAGTCT | 1862 |
| Db | 1300 | CAGGCTAAGATTTATATCGACTCTCTTTTGTACACACTGCACGCGCGCTTCTGCAGCCT | 1359 |
| Qy | 1863 | ATTGATGATCCATGACATACAGGCTCGACAGAGAGATAGCTTGGTCATATTTCTGAG | 1922 |
| Db | 1360 | ATCCAGATTCAGGACACACCGCGCCGACAGAGAGACAACTGGGTCACTTTCTGGAG | 1419 |
| Qy | 1923 | GAAATTTGCCACTTGCAGATGAGTTTATGACATTTTATTAATAGGCAGAGAGGTTG | 1982 |
| Db | 1420 | GAGTTGCCACACTTCAGGATG-----AGGACAGAGAGGTTG | 1456 |
| Qy | 1983 | ATGACGCGCTTTCACACCATGCTGTTTGAACAGGAAACCCCAAGGCAACATTTGGCTGTT | 2042 |
| Db | 1457 | ACGCTGCTTTTACGCTGCTGATGAAGCTGGAGCGCAGCGGAGCATTTAGCCTGTC | 1516 |
| Qy | 2043 | TAGGTACCTGGTCTTTACATTAACCTTGCATTTATGATGATGATGATGATGATGATGAT | 2102 |
| Db | 1517 | TGGGACCTGGATCCTTACCACACCTTGCATCATGATCCAGTACCTGCTCAGCGCT | 1576 |
| Qy | 2103 | TGAAATTCGAACTCTACAGATGACAGAGTACTATTACATATTTGGTATCTCTCTGAAT | 2162 |
| Db | 1577 | TTGAGTCTGAGCTCTACAGATGACAGAACTATTATCATCTACTGTTATCTGTGAGT | 1636 |
| Qy | 2163 | TCCTTTACGATGTTGATGTCAACTTGAAGTGTGCGGATGGCTCTCAAAATGGCAGAG | 2222 |
| Db | 1637 | TCTTGTATGCGTGGTGTGTCGAGTTCGAGTCGAGCGCACAGCTCTCAGATGGGAGG | 1696 |
| Qy | 2223 | AAAGATTAATGGAAGAGCAGCAGAAAGCGCTAGTGTAGTAAATAAACAAGAAAAAAGA | 2282 |
| Db | 1697 | AGAGAATCCTGGAGGAGCAGCTCAAAGTGCAGCAGCAAGAAAGAGCAAGAAAGAAGA | 1756 |
| Qy | 2283 | AAATTCGCGCATTTGAGCGCAGAGATCAAACTGAGCCAAAGCATATCAGAAATGTGTGCTG | 2342 |
| Db | 1757 | AACTCTGCTCTGAGTAAAGAGATCAACATGAGCCAGGCTTACAGAAACATGTGTGCTG | 1816 |
| Qy | 2343 | GAAATTTTAAACCATGTTAGATTTGACATGAGCGGCAAGTACGTAAACCGAAGTTTG | 2402 |
| Db | 1817 | GGATGTCAAGACGATGATCGCGTGGATATGGATCGAAAGGTGGCAAAACCTCAGTTTG | 1876 |
| Qy | 2403 | AGCTTTAGTGAACAAGTTCGGTATGAACACAGGTTTGTCTTCACTTCAACAGTGTGATGA | 2462 |
| Db | 1877 | AGCTGGACAGTGAAGGTTGCGTACGAGCATTCGCTTCCGCTTTCACAGTGTGCTCA | 1936 |
| Qy | 2463 | CCCCCGCCCGATGACACTTACTTACAGTTTCAAGGAAATGTCTGACCTCAATAATATAGCC | 2522 |
| Db | 1937 | CACACACACAGTTCATCATCTCAGTTTAAAGAAATGTCCGATTTGAAGATGACATC | 1996 |
| Qy | 2523 | CTCTCTCTCAGTCTCTGAACTGTATGTTGGCAGCTTAGTAAAGCATTTTCAACAGGCAAAA | 2582 |
| Db | 1997 | CTCTCTCTCTCAGCGGATCTCTATATGGCTGCCAGTAAACATTTCCAGCAGGCGCAAC | 2056 |
| Qy | 2583 | TGATATTCGAAAAATTTCTTAACCGGACCATGAGTTTAAATAGATTTTAAAGTTTGCA | 2642 |
| Db | 2057 | TCTTACTGAAAAATGTCACCGCCCTGATGAGGTTTAACTGATCTCTTAAAGTGGCCA | 2116 |
| Qy | 2643 | AAACCAACTTTTGTGTTTAAAGTTTATTTGGCAGGAGGACACAAAAAGGAATCTTAAAGTTT | 2702 |
| Db | 2117 | AAACCAATATTTGTTGTTTAAAGTCTGCTGGCTGGAGGACACAGAGGAGCAAGAGGCC | 2176 |

```

QY      2703 CTCTGAATTGATTTCTCTGCTCATATAATTTTCTGTTGTGGAACCTTGTGTTGAGA 2760
Db      |||||
2177 TTCTGAATTGATTTCTCTGCTCAATAACTTCCCCCTAGTCAAAATCCTATGAGA 2234
|||

RESULT 7
AK026296      997 bp      mRNA      linear      PRI 29-SEP-2000
LOCUS      Homo sapiens cDNA: FLJ22643 fis, clone HS107031.
DEFINITION      AK026296
ACCESSION      AK026296.1 GI:10439112
VERSION      oligo capping; fis (full insert sequence).
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (sites)
AUTHORS      Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 997)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES             Location/Qualifiers
     source           1..997
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HS107031"
                     /tissue_type="human small intestine"
                     /clone_lib="HSI"
                     /note="Cloning vector pME18SFL3"
     CDS              15..650
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB15435.1"
                     /db_xref="GI:10439113"
                     /translation="MHEYYIYVLSFLYAWLMSLISRADGSGMAEERIMERQOQR
SRKTKKKKVRLEIRITQSYQNNCAQFMKTVAFDMGDKVRPKFELDSQVRY
EHFAPNSVTPPPHYLQFKMSDLNKYSPFPQPELYVAASKHFOQAAMILENP
NPDVENVRLKVPKPNVVKMLAGGHKESKVPPEPFDFAHKYFPVKLV"
BASE COUNT      334 a 171 c 213 g 279 t
ORIGIN

Query Match      31.8%; Score 988; DB 9; Length 997;
Best Local Similarity 100.0%; Pred. No. 4.4e-213;
Matches 988; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2109 TGGAACTCTACAGTATGCACAGTACTATTACATATATTGGTATCTCTCTGAATTCCTTT 2168
Db      |||||
1 TGGAACTCTACAGTATGCACAGTACTATTACATATATTGGTATCTCTCTGAATTCCTTT 60
|||

QY      2169 ACCGATGTTGATGTCAACATTGAGTGTGCGGATGCTCTCAAAATGGCAGAGAAAGGA 2228
Db      |||||
61 ACCGATGTTGATGTCAACATTGAGTGTGCGGATGCTCTCAAAATGGCAGAGAAAGGA 120
|||

QY      2229 TAATGGNAGAGCAGCAAGGCGGTAGTAGTAAAAAACAAGAAAAAAGAAAGTTC 2288
|||

```

```

Db      |||||
121 TAATGGAAGACGACGAGAAAGCGGTAGTAGTAAAAAACAAGAAAAAAGAAAGTTC 180
|||
QY      2289 GCCAATTGAGCGGAGAGATCAAAATGAGCCAAAGCATATCAGAAATGTGTCTCTGGAATGT 2348
Db      |||||
181 GCCAATTGAGCGGAGAGATCAAAATGAGCCAAAGCATATCAGAAATGTGTCTCTGGAATGT 240
|||
QY      2349 TTAACCAATGAGTGTAGCATTTGACATGACGCGCAAGTACGTAAACCGAAGTTTCAGCTTG 2408
Db      |||||
241 TTAACCAATGAGTGTAGCATTTGACATGACGCGCAAGTACGTAAACCGAAGTTTCAGCTTG 300
|||
QY      2409 ATAGTGAACAAGTTTCGGTATGAAACACAGAGTTTGTCTCCATTCAACAGTGTGATACCCCGC 2468
Db      |||||
301 ATAGTGAACAAGTTTCGGTATGAAACACAGAGTTTGTCTCCATTCAACAGTGTGATACCCCGC 360
|||
QY      2469 CGCCAGTGCACTACTTACAGTTCAAGGAATGTCTGACCTCAATAATAATATAGCCCTCTC 2528
Db      |||||
361 CGCCAGTGCACTACTTACAGTTCAAGGAATGTCTGACCTCAATAATAATATAGCCCTCTC 420
|||
QY      2529 CTCAGTCTCTGAACTGTATGTGCGCAGCTAGTAGCACTTTCAACAGGCAAAAAATGATAT 2588
Db      |||||
421 CTCAGTCTCTGAACTGTATGTGCGCAGCTAGTAGCACTTTCAACAGGCAAAAAATGATAT 480
|||
QY      2589 TGGAAAAATATTCTCTAACCCGACCATGAGGTTAATAGAAATTTTAAAGTTGCCAAACCCA 2648
Db      |||||
481 TGGAAAAATATTCTCTAACCCGACCATGAGGTTAATAGAAATTTTAAAGTTGCCAAACCCA 540
|||
QY      2649 ACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTTAAAGTTCTCTCTG 2708
Db      |||||
541 ACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTTAAAGTTCTCTCTG 600
|||
QY      2709 AATTTGATTTCTCTGCTCATAAATATTTTCTCTGTGTGAAACTGTTTGGAGAGAGACTGG 2768
Db      |||||
601 AATTTGATTTCTCTGCTCATAAATATTTTCTCTGTGTGAAACTGTTTGGAGAGAGACTGG 660
|||
QY      2769 GGAGGTGCCCATAAAGGGGAGAGTCTTCTTTTTCAGACCCCACTCTAGAGGGCACATCAC 2828
Db      |||||
661 GGAGGTGCCCATAAAGGGGAGAGTCTTCTTTTTCAGACCCCACTCTAGAGGGCACATCAC 720
|||
QY      2829 CAGGCTCCACATCAGCGGAAGTGAGATGGATTTCTTGGGTAACTCACTATTATAGGAAT 2888
Db      |||||
721 CAGGCTCCACATCAGCGGAAGTGAGATGGATTTCTTGGGTAACTCACTATTATAGGAAT 780
|||
QY      2889 ACTTTTAGTTTGCAGCGCTTATATGACATGAATGAATAAAGTCTCTTTTAAAGTGGTTTAT 2948
Db      |||||
781 ACTTTTAGTTTGCAGCGCTTATATGACATGAATGAATAAAGTCTCTTTTAAAGTGGTTTAT 840
|||
QY      2949 TATGTTCCATGGAAGAACTGGTCTTATTGAATGCATTGATGAACGTTATATGTTTAT 3008
Db      |||||
841 TATGTTCCATGGAAGAACTGGTCTTATTGAATGCATTGATGAACGTTATATGTTTAT 900
|||
QY      3009 TACAGATTTAATCACAATCATTTTATTGAATGATTTGAGTGAATAATAGTGTTTATAAAG 3068
Db      |||||
901 TACAGATTTAATCACAATCATTTTATTGAATGATTTGAGTGAATAATAGTGTTTATAAAG 960
|||
QY      3069 GTTAAATAAATTTCTTGACAAAAAATAA 3096
Db      |||||
961 GTTAAATAAATTTCTTGACAAAAAATAA 988
|||

RESULT 8
AK056059      1796 bp      mRNA      linear      PRI 01-AUG-2002
LOCUS      Homo sapiens cDNA FLJ131497 fis, clone NT2NE2005317, highly similar
DEFINITION      to Rattus norvegicus corneal wound healing related protein mRNA.
ACCESSION      AK056059
VERSION      AK056059.1 GI:16551204
KEYWORDS      oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1

```

AUTHORS
 Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T., Kodaira, H., Furiya, T., Takahashi, M., Kikkawa, B., Omura, Y., Abe, K., Kamiyama, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Murakawa, K., Kaneshiro, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 1796)
 Isogai, T., Otsuki, T. and Sugiyama, T.
 Direct Submission
 Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - & 3' - end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.
 FEATURES
 Location/Qualifiers
 1..1796
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2NE2005317"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone_lib="NT2NE2"
 /note="Cloning vector: pWE18SFL3-mRNA from NT2 neuron after the differentiation of NT2 neuronal precursor cells."
 BASE COUNT 490 a 365 c 463 g 478 t
 ORIGIN
 Query Match 28.6%; Score 885.4; DB 9; Length 1796;
 Best Local Similarity 99.9%; Pred. No. 9e-150;
 Matches 886; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 549 AGGCATAATGGTTATGAAGCTTCTGTAGATGATGACGATTCAGGATGGAGCTCAGTAT 608
 DB 345 AGGCATAATGGTTATGAAGCTTCTGTAGATGATGACGATTCAGGATGGAGCTCAGTAT 404
 QY 609 GCCGAAAAATGGGAAGCAATACAACTGGGTGGACATTCACCAAGATTTTGAGA 668
 DB 405 GCCGAAAAATGGGAAGCAATACAACTGGGTGGACATTCACCAAGATTTTGAGA 464
 QY 669 AGCTTGTCGAGAAATTAAGTTGGGAGAACTACTTCATGATAGCTATTGGTCTTTTGA 728
 DB 465 AGCTTGTCGAGAAATTAAGTTGGGAGAACTACTTCATGATAGCTATTGGTCTTTTGA 524
 QY 729 AGCCATGTCGTATTGAAATGATGATGCCCAAGATGATGCTGCATGATTTGGAACCA 788
 DB 525 AGCCATGTCGTATTGAAATGATGATGCCCAAGATGATGCTGCATGATTTGGAACCA 584
 QY 789 AGTTAATCGAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAATTA 848
 DB 585 AGTTAATCGAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAATTA 644
 QY 849 AGATCTACCTTGCCCTGAACTGATAGGATTAAGGATACATGTTTGTGTTGATAAC 908
 DB 645 AGATCTACCTTGCCCTGAACTGATAGGATTAAGGATACATGTTTGTGTTGATAAC 704
 QY 909 GTGGTTAGAGGCCAATTCCTGGGACAGACAGTATTATTAGTGCCCTTTACATTCATA 968
 DB 705 GTGGTTAGAGGCCAATTCCTGGGACAGACAGTATTATTAGTGCCCTTTACATTCATA 764
 QY 969 AGACTTTATAGAAGATCCTGCTACTGAAGGCTTTTGTCTCTGGGAATCTTGAATAATCTGTGA 1028
 Db 765 AGACTTTATAGAAGATCCTGCTACTGAAGGCTTTTGTCTCTGGGAATCTTGAATAATCTGTGA 824
 QY 1029 CATTGCAAGGGAAGAAAGCTAAATAAGCTGCTGTTTGTGAGAGGAAGATTTTCAGTCAAT 1088
 Db 825 CATTGCAAGGGAAGAAAGCTAAATAAGCTGCTGTTTGTGAGAGGAAGATTTTCAGTCAAT 884
 QY 1089 GACTTATGATTTTAAAAATGGCTAACAGTGTGACAGATCTTTCGAGTTACAGGCATGCTAAA 1148
 Db 885 GACTTATGATTTTAAAAATGGCTAACAGTGTGACAGATCTTTCGAGTTACAGGCATGCTAAA 944
 QY 1149 AGATGTGAGAGTACATGCAAGAAAGAGTAAAGAGTACTCGAAGTCGACAGAGAGA 1208
 Db 945 AGATGTGAGAGTACATGCAAGAAAGAGTAAAGAGTACTCGAAGTCGACAGAGAGA 1004
 QY 1209 AAGAGATCCGAAAGTTCGAACCTAGAACCAACCAATGTTTAGCAGTATTTCAGCAGGTGAA 1268
 Db 1005 AAGAGATCCGAAAGTTCGAACCTAGAACCAACCAATGTTTAGCAGTATTTCAGCAGGTGAA 1064
 QY 1269 ATTTACTCGTGTGTTACTGACAGTCTTATAGCCTTTTACTAAGAAAGAGACAGGTGCTGT 1328
 Db 1065 ATTTACTCGTGTGTTACTGACAGTCTTATAGCCTTTTACTAAGAAAGAGACAGGTGCTGT 1124
 QY 1329 TGCAGAGCTCAAAAATGATGTTCAAGCAGCAGATCTTCTTCTGCCATTCATAATTC 1388
 Db 1125 TGCAGAGCTCAAAAATGATGTTCAAGCAGCAGATCTTCTTCTGCCATTCATAATTC 1184
 QY 1389 ATTGCATCATGGCATCCAGGCCAGAAATGATATACAAAAAGAGATC 1435
 Db 1185 ATTGCATCATGGCATCCAGGCCAGAAATGATATACAAAAAGAGATC 1231
 RESULT 9
 AX664135 LOCUS 719 bp DNA linear PAT 22-MAR-2003
 DEFINITION Sequence 7 from Patent WO02057304.
 ACCESSION AX664135
 VERSION AX664135.1 GI:29164145
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 Panzer, S.R., Lincoln, S.E., Altus, C.M., Dufour, G.E., Jackson, J.L., Jones, A.L., Dam, T.C., Liu, T.F., Harris, B., Flores, V., Daffo, A., Marwaha, R., Chen, A.J., Chang, S.C., Gerstein, E.H., Peralta, C.H., David, M.H., and Lewis, S.A.
 Secretory molecules
 Patent: WO 02057304-A 7 25-JUL-2002;
 Incyte Genomics, Inc. (US)
 Location/Qualifiers
 source 1..719
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /note="Incyte ID No: LI:462889.1:2001JUN12"
 BASE COUNT 223 a 113 c 188 g 195 t
 ORIGIN
 Query Match 23.2%; Score 717.4; DB 6; Length 719;
 Best Local Similarity 99.9%; Pred. No. 9.9e-152;
 Matches 718; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 474 GCGCGCGCGCGAGCGCGCTGTTATTTCCCGTCCGACAGTCGCTGGCGCGCGCGGT 533
 Db 1 GCGCGCGCGCGAGCGCGCTGTTATTTCCCGTCCGACAGTCGCTGGCGCGCGGT 60
 QY 534 GACCAACCGGAGAGTAGGCATAATGGTTATGAAAGCTTTCTGTAGATGATGATGATCAGG 593
 Db 61 GACCAACCGGAGAGTAGGCATAATGGTTATGAAAGCTTTCTGTAGATGATGATGATCAGG 120
 QY 594 ATGGAGGCTCAGTATGCCAGAAAAAATCGAGAAAGCAATACAACTGGGTGGACATTAC 653

| | | | |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|----------------------------|
| Db | 121 | ATGGGAGCTCAGTATGCGCAGAAAAATGGAGAAAGCAATACAAACTGGGTGGACATTAC | 180 |
| Qy | 654 | CCAGAGATTTTGAAGAAGCTTGTGAGAGATTAAGCTTGGGAGAACTACTTCATGATAAGCT | 713 |
| Db | 181 | CCAGAGATTTTGAAGAAGCTTGTGAGAGATTAAGCTTGGGAGAACTACTTCATGATAAGCT | 240 |
| Qy | 714 | ATTGTGCTCTTTTGAAGCCATGCTGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGG | 773 |
| Db | 241 | ATTGTGCTCTTTTGAAGCCATGCTGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGG | 300 |
| Qy | 774 | CATGATTTGGAACCAAGTAAATCAAAAAGTCTCAATTTTGAACAGCTATCAAGGATGG | 833 |
| Db | 301 | CATGATTTGGAACCAAGTAAATCAAAAAGTCTCAATTTTGAACAGCTATCAAGGATGG | 360 |
| Qy | 834 | CACATTTAAATTAAGATCTCACCTTGCCTGAACTGATAGGATTTATGGATACATGTTT | 893 |
| Db | 361 | CACATTTAAATTAAGATCTCACCTTGCCTGAACTGATAGGATTTATGGATACATGTTT | 420 |
| Qy | 894 | TTGCTGTTTGATAACGTGGTTAGAGGCCATTCCTACCTGCGCACAGACAGTATTTCAGTGCCT | 953 |
| Db | 421 | TTGCTGTTTGATAACGTGGTTAGAGGCCATTCCTACCTGCGCACAGACAGTATTTCAGTGCCT | 480 |
| Qy | 954 | TTACATTCATAATCCAGACTTTATAGAGATCTCTGCTATGAGAGCTTTTGGCTCTGGGAAT | 1013 |
| Db | 481 | TTACATTCATAATCCAGACTTTATAGAGATCTCTGCTATGAGAGCTTTTGGCTCTGGGAAT | 540 |
| Qy | 1014 | CTTGAATAATCTGTGACATTTGCAAGGGAAAAAGTAAATAAGCTGCTGTTTTGAAGAGGA | 1073 |
| Db | 541 | CTTGAATAATCTGTGACATTTGCAAGGGAAAAAGTAAATAAGCTGCTGTTTTGAAGAGGA | 600 |
| Qy | 1074 | AGATTTTCAGTCAATGATCTTATGAGATTTAAATGGCTAAACAGTGTGACAGATCTTCGAGT | 1133 |
| Db | 601 | AGATTTTCAGTCAATGATCTTATGAGATTTAAATGGCTAAACAGTGTGACAGATCTTCGAGT | 660 |
| Qy | 1134 | TACAGGATCTCTAAAGATGTGGAGGATGACATCCCAAGAGAGATAAGATCTCGAA | 1192 |
| Db | 661 | TACAGGATCTCTAAAGATGTGGAGGATGACATCCCAAGAGAGATAAGATCTCGAA | 719 |
| RESULT 10 | | | |
| AC102010 | | | |
| LOCUS | AC102010 | 147876 bp | DNA linear HTG 21-MAR-2003 |
| DEFINITION | Mus musculus clone RP24-531M24, WORKING DRAFT SEQUENCE, 7 unordered | | |
| ACCESSION | AC102010 | | |
| VERSION | AC102010.3 | GI:29135693 | |
| KEYWORDS | HTG; HTGS PHASE1; HTGS DRAFT. | | |
| SOURCE | Mus musculus (house mouse) | | |
| ORGANISM | Mus musculus | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| AUTHORS | Birren, B., Nusbaum, C. and Lander, E. | | |
| TITLE | Mus musculus, clone RP24-531M24 | | |
| JOURNAL | Unpublished | | |
| REFERENCE | 2. (bases 1 to 147876) | | |
| AUTHORS | Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chararo, B., Choepel, V., Collangelo, M., Collins, S., Collymore, A., Cooke, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glende, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Wu, X., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M. | | |

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 147876)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cooke, A., Cooke, P., Corum, B., DeArelano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hagos, B., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ranasingh, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (21-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 21, 2003 this sequence version replaced gi:28412092.

All repeats were identified using RepeatMasker:

Smith, A. F. A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L17873

Center clone name: 531_M.24

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 146099 bases at least Q40

Consensus quality: 146869 bases at least Q30

Consensus quality: 147004 bases at least Q20

Insert size: 157000; agarose-fp

Insert size: 147276; sum-of-contigs

Quality coverage: 9.9 in Q20 bases; agarose-fp

Quality coverage: 10.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 84958: contig of 84958 bp in length

* 84959

* 85059 87411: contig of 2353 bp in length

* 87412 87511: gap of 100 bp

* 87512 89933: contig of 2422 bp in length

* 89934 90036: gap of 100 bp

* 90034 94076: contig of 4043 bp in length

* 94077 94176: gap of 100 bp


```

sequence.
AL353743
VERSION
AL353743.22 GI:15552931
SOURCE
HTG.
ORGANISM
Homo sapiens (human)
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 181864)
Corby,N.
Direct Submission
Submitted (08-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 11, 2001 this sequence version replaced gi:14267784.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30) : an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: 'EMBL'; Sw: '
SWISSPROT'; Tr: 'TREMBL'; Wp: 'WORMPEP'; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-213G2 is from the library RPC1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VVECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-213G2 It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true right end of clone RP11-213G2 is at 181864 in this
sequence. The true left end of clone RP11-65C15 is at 111320 in
this sequence. The true right end of clone RP11-20211 is at 100 in
this sequence.
FEATURES
source
Location/Qualifiers
1..181864
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-213G2"
/clone_lib="RPC1-11.1"
39727
unsure
115957..116075
misc_feature
/note="Sequence from overlapping clone bA65C15 (AL161453)."
Assembly confirmed by restriction digest."
117488..117564
misc_feature
/note="Sequence from overlapping clone bA65C15 (AL161453)."
Assembly confirmed by restriction digest."
BASE COUNT 53151 a 36268 c 36336 g 56109 t
ORIGIN
Query Match 17.8%; Score 551.4; DB 9; Length 181864;
Best Local Similarity 99.8%; Pred. No. 1.1e-113;
Matches 552; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTCTCAGCAAACTCCAGGCGGTGTATAGCAACATAAATCGTTGTTCAGGCAGCAGTA 60
Db 174760 TTCTCAGCAAACTCCAGGCGGTGTATAGCAACATAAATCGTTGTTCAGGCAGTA 1748

```


ACCESSION AJ334413
VERSION AJ334413.1 GI:15878831
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 743)
AUTHORS Kutsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F., Kvasha, S.M., Podowski, R.M., Matushkin, Y.G., Gyanchandani, A., Muravenko, O.V., Levitsky, V.G., Kolchanov, N.A., Protopopov, A.I., Kashuba, V.I., Kissilev, L.L., Wasserman, W., Wahlestedt, C., and Zabarovsky, E.R.
TITLE NotI flanking sequences: a tool for gene discovery and verification of the human genome
JOURNAL Nucleic Acids Res. 30 (14), 3163-3170 (2002)
MEDLINE 22131767
PUBMED 12136098
REFERENCE 2 (bases 1 to 743)
AUTHORS Zabarovsky, E.R.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre, Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77, Sweden
FEATURES
source Location/Qualifiers
1. 743
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="NR1-ND24R"
BASE COUNT 219 a 154 c 166 g 203 t 1 others
ORIGIN
Query Match 9.7%; Score 299; DB 9; Length 743;
Best Local Similarity 95.4%; Pred. No. 7.9e-57;
Matches 308; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 1 TTCTCTACGAACTCCAGGCGCTGTATAGAAACATAAATCCGTTGTGAGGACAGTA 60
Db 323 TTCTCTACGAACTCCAGGCGCTGTATAGAAACATAAATCCGTTGTGAGGACAGTA 264
Qy 61 GCACGCTGTTGCTCTCGAGCTTGGCTGCTCGTTCGTCGCAACCACTAAGTCTACG 120
Db 263 GCACGCTGTTGCTCTCGAGCTTGGCTGCTCGTTCGTCGCAACCACTAAGTCTACG 204
Qy 121 CAAACCTCCACGGTTTCCTTCGCGCTTCGGCTCACCTTTCTAAGAAATTCCTGAGGAGCA 180
Db 203 CAAACCTCCAGGTTTCCTTCGCGCTTCGGCTCACCTTTTAAGAAATTCCTGAGGAGCA 144
Qy 181 GCGCAGACGGGGGGGCTCTGAGACTCCGGGCTCCGCTCTTTCCGGAAACCGCCCACTA 240
Db 143 GCGCAGACGGGGGGGCTCTGAGACTCCGGGCTCCGCTCTTTCCGGAAACCGCCCACTA 84
Qy 241 CCAGGACTCCGACAGAGGGTGAAAAAGATTAATTCGGTCTCGGATCGTCTTAATC 300
Db 83 CTAGGACTCCGTCAGAGAAAAAACCAGATTACTTCGGTCTCGGATCGTCTTAATC 24
Qy 301 TCGCAGAGAGAGAGCGCGCGC 323
Db 23 TCGCAGAGAGAGAGCGCGCGC 1

RESULT 15
HSA342328
LOCUS HSA342328 642 bp DNA linear PRI 18-JUL-2002
DEFINITION Homo sapiens genomic sequence surrounding NotI site, clone NR1-ND24C.
ACCESSION AJ342328
VERSION AJ342328.1 GI:15886745
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 642)
AUTHORS Kutsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F., Kvasha, S.M., Podowski, R.M., Matushkin, Y.G., Gyanchandani, A., Muravenko, O.V., Levitsky, V.G., Kolchanov, N.A., Protopopov, A.I., Kashuba, V.I., Kissilev, L.L., Wasserman, W., Wahlestedt, C., and Zabarovsky, E.R.
TITLE NotI flanking sequences: a tool for gene discovery and verification of the human genome
JOURNAL Nucleic Acids Res. 30 (14), 3163-3170 (2002)
MEDLINE 22131767
PUBMED 12136098
REFERENCE 2 (bases 1 to 642)
AUTHORS Zabarovsky, E.R.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre, Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77, Sweden
FEATURES
source Location/Qualifiers
1. 642
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="NR1-ND24C"
BASE COUNT 73 a 209 c 265 g 93 t 2 others
ORIGIN
Query Match 7.6%; Score 236.4; DB 9; Length 642;
Best Local Similarity 99.6%; Pred. No. 1.2e-42;
Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 316 GCGGCCGCCATCGGCCCGAACGAGCGGTGCGAGGAGGGGGTGTGGCCCGGAGCGCG 375
Db 1 GCGGCCGCCATCGGCCCGAACGAGCGGTGCGAGGAGGGGGTGTGGCCCGGAGCGCG 60
Qy 376 AAGTCCCGGAGTAAGGAGAGGGCGCGGTGCGCGTCCCGGCATACGATGCGTG 435
Db 61 AAGTCCCGGAGTAAGGAGAGGGCGCGGTGCGCGTCCCGGCATACGATGCGTG 120
Qy 436 CACGCTCCCGTGGGCTGGGCTGAGAGGGAGGGGGCGCGCGCGCGCGCGCGCG 495
Db 121 CACGCTCCCGTGGGCTGGGCTGAGAGGGAGGGGGCGCGCGCGCGCGCGCG 180
Qy 496 TTATTTCCGTGTCGCGACAGTGGTGGCGCGCGCGGTGACCCAGGAGAAATAGGCA 553
Db 181 TTATTTCCGTGTCGCGACAGTGGTGGCGCGCGGTGACCCAGGAGAAATAGGTA 238
Search completed: November 25, 2003, 01:06:53
Job time : 11149 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 25, 2003, 03:13:51 ; Search time 106 Seconds

(without alignments)
10683.169 Million cell updates/sec

Title: US-10-001-857-42

Perfect score: 5439

Sequence: 1 tctctcagcaaacctccagg.....atttttgacaaaaaaa 3096

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 667484 seqs, 182883703 residues

Total number of hits satisfying chosen parameters: 1334968

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlh

-Q=/cgn2_1/USFTC_spool/US10001857/runat_24112003_162856_9623/app_query.fasta_1.3271

-DB=PublishedApplications_AA -GFWT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1

-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62

-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100

-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0

-MAXLEN=2000000000 -USER=US10001857 @CGN 1.1 149 @runat_24112003_162856_9623

-NCPU=6 -ICPU=3 -NO MWAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100

-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*

6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*

7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*

9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp.*

10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp.*

11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp.*

12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*

13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp.*

14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp.*

15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp.*

16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*

17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*

18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| ----- | | | | | |

| 1 | 3796 | 69.8 | 733 | 14 | US-10-001-857-145 | Sequence 145, Appl |
|----|-------|------|------|----|---------------------|--------------------|
| 2 | 165 | 3.0 | 523 | 12 | US-10-017-161-1982 | Sequence 1982, Ap |
| 3 | 146.5 | 2.7 | 574 | 12 | US-10-168-097A-76 | Sequence 76, Appl |
| 4 | 146.5 | 2.7 | 574 | 12 | US-10-239-431A-38 | Sequence 38, Appl |
| 5 | 145 | 2.6 | 559 | 10 | US-09-858-155A-2 | Sequence 2, Appl |
| 6 | 145 | 2.6 | 559 | 12 | US-10-168-097A-26 | Sequence 26, Appl |
| 7 | 145 | 2.6 | 559 | 12 | US-10-239-431A-33 | Sequence 33, Appl |
| 8 | 145 | 2.6 | 559 | 12 | US-10-116-370-2 | Sequence 2, Appl |
| 9 | 144 | 2.6 | 1134 | 14 | US-10-001-873-50 | Sequence 50, Appl |
| 10 | 142.5 | 2.6 | 800 | 12 | US-10-029-386-32198 | Sequence 32198, A |
| 11 | 142 | 2.6 | 721 | 14 | US-10-086-464-5 | Sequence 5, Appl |
| 12 | 141.5 | 2.6 | 5877 | 15 | US-10-142-515-11 | Sequence 11, Appl |
| 13 | 141.5 | 2.6 | 5935 | 15 | US-10-243-431A-8 | Sequence 8, Appl |
| 14 | 140 | 2.5 | 802 | 9 | US-09-823-240-2 | Sequence 2, Appl |
| 15 | 139.5 | 2.5 | 1243 | 15 | US-10-196-935A-4 | Sequence 4, Appl |
| 16 | 138.5 | 2.5 | 641 | 11 | US-09-959-987-2 | Sequence 2, Appl |
| 17 | 138.5 | 2.5 | 647 | 9 | US-09-925-299-1002 | Sequence 1002, Ap |
| 18 | 138.5 | 2.5 | 647 | 11 | US-09-925-299-1002 | Sequence 2, Appl |
| 19 | 136.5 | 2.5 | 647 | 14 | US-10-086-464-2 | Sequence 4, Appl |
| 20 | 136.5 | 2.5 | 647 | 14 | US-10-086-464-4 | Sequence 4, Appl |
| 21 | 134.5 | 2.4 | 520 | 12 | US-10-168-097A-36 | Sequence 36, Appl |
| 22 | 134.5 | 2.4 | 520 | 12 | US-10-239-431A-34 | Sequence 34, Appl |
| 23 | 134 | 2.4 | 351 | 12 | US-10-239-431A-22 | Sequence 22, Appl |
| 24 | 134 | 2.4 | 542 | 12 | US-10-239-431A-10 | Sequence 10, Appl |
| 25 | 133.5 | 2.4 | 507 | 14 | US-10-078-547-24 | Sequence 24, Appl |
| 26 | 132.5 | 2.4 | 206 | 15 | US-10-156-761-14830 | Sequence 14830, A |
| 27 | 132.5 | 2.4 | 797 | 15 | US-10-156-761-10907 | Sequence 10907, A |
| 28 | 132 | 2.4 | 503 | 14 | US-10-078-547-2 | Sequence 2, Appl |
| 29 | 131 | 2.4 | 528 | 12 | US-09-840-746-20 | Sequence 20, Appl |
| 30 | 131 | 2.4 | 538 | 10 | US-09-976-740-43 | Sequence 43, Appl |
| 31 | 131 | 2.4 | 538 | 14 | US-10-023-529-43 | Sequence 43, Appl |
| 32 | 131 | 2.4 | 538 | 14 | US-10-023-523-43 | Sequence 43, Appl |
| 33 | 131 | 2.4 | 235 | 15 | US-10-050-704-102 | Sequence 102, App |
| 34 | 130 | 2.4 | 455 | 12 | US-10-032-585-7399 | Sequence 7399, Ap |
| 35 | 129.5 | 2.4 | 501 | 12 | US-10-168-097A-46 | Sequence 46, Appl |
| 36 | 129.5 | 2.4 | 501 | 12 | US-10-239-431A-35 | Sequence 35, Appl |
| 37 | 129 | 2.3 | 377 | 15 | US-10-149-819-7 | Sequence 7, Appl |
| 38 | 129 | 2.3 | 505 | 12 | US-10-168-097A-56 | Sequence 56, Appl |
| 39 | 129 | 2.3 | 405 | 12 | US-10-239-431A-36 | Sequence 36, Appl |
| 40 | 126.5 | 2.3 | 405 | 12 | US-10-017-161-2256 | Sequence 2256, Ap |
| 41 | 126.5 | 2.3 | 1046 | 15 | US-10-156-761-10088 | Sequence 10088, A |
| 42 | 126 | 2.3 | 416 | 15 | US-10-043-487-282 | Sequence 282, App |
| 43 | 126 | 2.3 | 505 | 12 | US-10-168-097A-12 | Sequence 12, Appl |
| 44 | 126 | 2.3 | 505 | 12 | US-10-239-431A-32 | Sequence 32, Appl |
| 45 | 125 | 2.3 | 550 | 10 | US-09-976-740-47 | Sequence 47, Appl |

ALIGNMENTS

RESULT 1

US-10-001-857-145
; Sequence 145, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 145
; LENGTH: 733
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-857-145

Alignment Scores:

Pred. No.: 5e-294 Length: 733
Score: 3796.00 Matches: 732
Percent Similarity: 99.59% Conservative: 0
Best Local Similarity: 99.58% Mismatches: 1
Query Match: 69.79% Indels: 3
DB: 14 Gaps: 0

US-10-001-857-42 (1-3096) x US-10-001-857-145 (1-733)

| | | | |
|----|------|-----------------------------------------------------------------|------|
| QY | 556 | ATGGTTATGAAGCTCTGTAGTATGATGACGATTCAGGATGGAGCTCAGTATGCCAGAA | 615 |
| DB | 1 | MetValMetLysAlaSerValAspAspAspSerGlyTrpGluLeuSerMetProGlu | 20 |
| QY | 616 | AAATGGAGAAAAGCAATACAACTGGGTGGACATTTACCCCAAGATTTTGAAGAAGCTTGT | 675 |
| DB | 21 | LysMetGluLysSerAsnThrAsnTrpValAspIleThrGlnAspPheGluGluAlaCys | 40 |
| QY | 676 | CGAAGATTAAAGTTGGGAACTACTTCATGTATAGCTATTGGTCTTTTGAAGCCATG | 735 |
| DB | 41 | ArgGluLeuLysLeuGlyGluLeuLeuHisAspLysLeuPheGlyLeuPheGluAlaMet | 60 |
| QY | 736 | TCTGCTATTGAATGATGCCATCCCAAGATGGATGCTGGCATGATTGGAAACCAAGTTAAT | 795 |
| DB | 61 | SerAlaIleGluMetMetAspProLysMetAspAlaGlyMetIleGlyAsnGlnValAsn | 80 |
| QY | 796 | CGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAAAATTAAAGATCTC | 855 |
| DB | 81 | ArgLysValLeuAsnPheGluGlnAlaIleLysAspGlyThrIleLysIleLysAspLeu | 100 |
| QY | 856 | ACCTTGCCCTGAAGTATAGGATTATGGATACATGTTTTGCTGTTGTATAGCTGGTTA | 915 |
| DB | 101 | ThrLeuProGluLeuIleGlyIleMetAspThrCysPheCysLeuLeuIleThrTrpLeu | 120 |
| QY | 916 | GAAAGGCATTCACTGGCAGACAGATATTACGTGCTTTACATTCATAATCAGACTTT | 975 |
| DB | 121 | GluGlyHisSerLeuAlaGlnThrValPheThrCysLeuTyrlleHisAsnProAspPhe | 140 |
| QY | 976 | ATAGAGATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAATACTGTGACATTGCA | 1035 |
| DB | 141 | IleGluAspProAlaMetLysAlaPheAlaLeuGlyIleLeuLysIleCysAspIleAla | 160 |
| QY | 1036 | AGGAAAAGTAAATAAAGCTGCTGTTTTGAGAGGAGAGATTTTCAGTCAATGACTTAT | 1095 |
| DB | 161 | ArgGluLysValAsnLysAlaAlaValPheGluGluGluAspPheGlnSerMetThrTrp | 180 |
| QY | 1096 | GGATTTTAAATGGCTTAACAGTGTGACAGATCTTTCGAGTTACAGGCATGCTTAAAGATGTG | 1155 |
| DB | 181 | GlyPheLysMetAlaAsnSerValThrAspLeuArgValThrGlyMetLeuLysAspVal | 200 |
| QY | 1156 | GAGGATGACATGCAAGAAGAGTAAAGAGTACTCGAAGTCGACAGAGAGAAAGAGAT | 1215 |
| DB | 201 | GluAspAspMetGlnArgValLysSerThrArgSerArgGlnGlyGluGluArgAsp | 220 |
| QY | 1216 | CCAGAAGTTGAATAGAACACCAACAAATGTTTAGCTAGTATTACAGCAGTGAATTTACT | 1275 |
| DB | 221 | ProGluValGluLeuGluHisGlnGlnCysLeuAlaValPheSerArgValLysPheThr | 240 |
| QY | 1276 | CGTGTGTTACTGACAGTCTTATAGCTTTTACTAAGAAGAGACCAAGTCTGTGCGAA | 1335 |
| DB | 241 | ArgValLeuLeuThrValLeuIleAlaPheThrLysLysLysGluThrSerAlaValAlaGlu | 260 |
| QY | 1336 | GCTCAAAAATTGATGTTCAAGCAGCAGATCTTCTTCTGCCATTTCAATTCATTGCGAT | 1395 |
| DB | 261 | AlaGlnLysLeuMetValGlnAlaAlaAspLeuLeuSerAlaIleHisAsnSerLeuHis | 280 |
| QY | 1396 | CATGGCATCCAGCCCGACAGATGATCTACAAAAGGAGATCATCAATTTATGATGGTTTT | 1455 |
| DB | 281 | HisGlyIleGlnAlaGlnAsnAspThrThrLysGlyAspHisProIleMetMetGlyPhe | 300 |
| QY | 1456 | GAACCCCTTGTGAACACAGGCTACTTCCACCTTACCTTCCCTCGATATGCAAAAATAAT | 1515 |
| DB | 301 | GluProLeuValAsnGlnArgLeuLeuProProThrPheProArgTyrlleAlaLysIleIle | 320 |

| | | | |
|----|------|----------------------------------------------------------------|------|
| QY | 1516 | AAAAGGGAAGAAATGTTGAACCTATTATTTGCAAGATTAAATAGATAGATAAAAACTGTCT | 1575 |
| DB | 321 | LysArgGluGluMetValAsnTyrPheAlaArgLeuIleAspArgIleLysThrValCys | 340 |
| QY | 1576 | GAGGTTTGAATTTAAACAAATTTACATTTGATCCTCGATTTTCTGTGAATTTAGTGAA | 1635 |
| DB | 341 | GluValValAsnLeuThrAsnLeuHisCysIleLeuAspPhePheCysGluPheSerGlu | 360 |
| QY | 1636 | CAGTCACCATGTGTTCTTTCAAGATCTCTGTATACAAACACATTTCTGTGGATACAAA | 1695 |
| DB | 361 | GlnSerProCysValLeuSerArgSerLeuLeuGlnThrThrPheLeuValAspAsnLys | 380 |
| QY | 1696 | AAGGCTTTTGGAACTCATCTCATGCAAGCATGTGTGAAAGATGCATTCGGTCTTTTGTG | 1755 |
| DB | 381 | LysValPheGlyThrHisLeuMetGlnAspMetValLysAspAlaLeuArgSerPheVal | 400 |
| QY | 1756 | AGATCCTCGAGTGTCTTCCCAAGTGTACTATATAATATACACAGGCTTAAGGACT | 1815 |
| DB | 401 | -AspProPro-ValLeuSerProLysCysTyrLeuTyAsnAsnHisGlnAlaLysAspC | 420 |
| QY | 1816 | GTATCGACTCTCTTTCTACTCTACTGTGTTGGCCATTTCTGTAGTCTTATTTCAGATCCATG | 1875 |
| DB | 420 | ysIleAspSerPheValThrHisCysValArgProPheCysSerLeuIleGlnIleHisG | 440 |
| QY | 1876 | GACATAACAGGGCTCGACAGAGATAGCTTGTGTCATATTCTTGAGGAATTTGCCACCT | 1935 |
| DB | 440 | lyHisAsnArgAlaargGlnArgAspLysLeuGlyHisIleLeuGluGluPheAlaThrL | 460 |
| QY | 1936 | TGCAGATGAGTTATGACATTTTATTTAAATAGCA-CAGAAAGGTTGTATGAGCGCTTC | 1994 |
| DB | 460 | euglnAspGluPheMetThrPheTyrPheAsnArgAlaGlnLysValAspAlaIleLeuH | 480 |
| QY | 1995 | ACACCATGCTGTGAAACACAGGAACCCCAAGGCAACATTTGGCTCTTTAGGTACTCTGGG | 2054 |
| DB | 480 | isthMetLeuLeuLysGlnGluProGlnArgGlnHisLeuAlaCysLeuGlyThrTrpV | 500 |
| QY | 2055 | TCCTTTACATAACCTTCGCATTTATGATACAGTACCTTCTTAAGTGGCTTTTGAATTGGAAC | 2114 |
| DB | 500 | alLeuTyrHisAsnLeuArgIleMetIleGlnTyrLeuLeuSerGlyPheGluLeuGluL | 520 |
| QY | 2115 | TCTACAGTATGACAGTACTATTACATATATTGTTATCTCTCTGAATTCCTTTACGCAT | 2174 |
| DB | 520 | eutySerMetHisGluTyrTyrIleTyrTrpTyrLeuSerGluPheLeuTyrlaIat | 540 |
| QY | 2175 | GGTTGATGTCAACATTGAGTCGTGGCATGGCTCTCAAAATGGCAGAGGAAGATATGG | 2234 |
| DB | 540 | rpLeuMetSerThrLeuSerArgAlaAspGlySerGlnMetAlaGluGluArgIleMetG | 560 |
| QY | 2235 | AAAGCAGCAGAAAGCCCGTAGTAGTAAAAAACAAGAAAAAAGAAAGTTGCGCCAT | 2294 |
| DB | 560 | luGluGlnGlnLysGlyArgSerSerLysLysThrLysLysLysLysValArgProL | 580 |
| QY | 2295 | TGAGCCGAGAGATCAATGAGCCAGCATATCAGAACATGTGTGTGGAATCTTTAAAA | 2354 |
| DB | 580 | eusSerArgIleThrMetSerGlnAlaTyrGlnAsnMetCysAlaGlyMetPheLysT | 600 |
| QY | 2355 | CCATGCTAGCATTTGACATGCGCAAGTACGTAAACCGAAGTTTGAGCTTTGATGTG | 2414 |
| DB | 600 | hrMetValAlaPheAspMetAspGlyLysValArgLysProLysPheGluLeuAspSerG | 620 |
| QY | 2415 | AACAGTTTCGGTATGACACAGGTTTGTCTCCATTCAACAGTGTGATGACCCGCCCGCAG | 2474 |
| DB | 620 | luGlnValArgTyrGluHisArgPheAlaProPheAsnSerValMetThrProProPro | 640 |
| QY | 2475 | TGCACCTACTTACAGTTCAAGGAAATGCTGACCTCAATAATAATATAGCCCTCTCTCTCAGT | 2534 |
| DB | 640 | alHisTyrLeuGlnPheLysGluMetSerAspLeuAsnLysTyrSerProProProGlnS | 660 |
| QY | 2535 | CTCCTGAACTGTATGTGGCAGCTAGTAGCATTTCACAGGCAAAATATGATTTGGAAA | 2594 |
| DB | 660 | erProGluLeuTyrValAlaAlaSerLysHisPheGlnGlnAlaLysMetIleLeuGluAl | 680 |

| | | | |
|------------------------|--------|---------------|----|
| Percent Similarity: | 40.52% | Conservative: | 14 |
| Best Local Similarity: | 31.37% | Mismatches: | 61 |
| Query Match: | 2.64% | Indels: | 30 |
| DB: | 10 | Gaps: | 8 |

US-10-001-857-42 (1-3096) x US-09-858-155A-2 (1-559)

| | | | |
|-----|-----|--------------------------------------------------------------|-----|
| 559 | QY | CCATTATGCGCTA-----CCTTCCTCCGTGGTGACCCGCGGCCGCAGCGCACT | 515 |
| | | ::: | |
| 294 | Db | ProThrCysIleSerSerAlaThrGlyLeuIleGluAsnArgProGlnSerProAlaThr | 313 |
| | | ::: | |
| 514 | QY | GTCGCGACCACGGAAATAACAGACGCGCTCGGCGCGCGCCCCCTCCCTCTCAGCC | 455 |
| | | :: | ::: |
| 314 | Db | GlyargThrProValpheValserProThrProProProProProProLeuProSer | 333 |
| | | :: | ::: |
| 454 | QY | CAGCCCAGCCGACGCGTGCACGCATGCGTATGCCCGCGGACGGCACCCC-----GCC | 401 |
| | | | |
| 334 | Db | AlaLeuSerThrSerSerLeuArgAlaSerMetThrSerThrProProProValPro | 353 |
| | | | |
| 400 | QY | CCCTCTCCCTTACTCCCGGGACITCG---CGCTCCCGCGGCACACCCCTC----- | 353 |
| | | ::: | |
| 354 | Db | ProProProProProAlaThrAlaLeuGlnAlaProAlaValProProProAla | 373 |
| | | ::: | |
| 352 | QY | -----CCTCGCCACCG-----CCTCGCTTC | 332 |
| | | ::: | |
| 374 | Db | ProLeuGlnIleAlaProGlyValLeuHisProAlaProProProIleAlaProLeu | 393 |
| | | | |
| 331 | QY | GGCCGATGGCGCGCCTCTCTCTCCGAGATTAGACGATCGGAGACCGGAAGTT | 272 |
| | ::: | :: ::: | |
| 394 | Db | ValGlnProSerProPro-----ValAlaArgAlaAlaProValCys-----GluThr | 409 |
| | | :: | |
| 271 | QY | ATCTTTTTTACCCTCTGTGCGAGTCCTGGGTAGTGGCGGTTCCTCCGGAAGAGCGGGAG | 212 |
| | | ::: | |
| 410 | Db | ValProValHisProLeuProGlnGlyValGlnGlyLeuProProProProPro | 429 |
| | | | |
| 211 | QY | CCCGGAGTCTCAGACCGCGCCCCCGTCTCGCTGCGCCTCT | 173 |
| | | :: | |
| 430 | Db | ProProLeu-----ProProGlyIleArqProSer | 440 |
| | | :: | |

RESULT 6

```

US-10-168-097A-26
; Sequence 26, Application US/10168097A
; Publication NO. US20030166245A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL OF THE RECHERCHE SCIENTIFIQUE
; APPLICANT: INSTITUT CURIE
; TITLE OF INVENTION: WASP FAMILY PROTEIN FRAGMENTS, AND USES THEREOF
; FILE REFERENCE: IFB99WASP
; CURRENT APPLICATION NUMBER: US/10/168,097A
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 559
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-168-097A-26

```

| | |
|------------------------|---------|
| Alignment Scores: | |
| Pred. No.: | 0.00942 |
| Score: | 145.00 |
| Length: | 48 |
| Matches: | 40.53% |
| Percent Similarity: | 31.3% |
| Best Local Similarity: | 31.3% |
| Mismatches: | 61 |
| Indels: | 30 |
| Query Match: | 2.64% |
| DB: | 12 |
| Gaps: | 8 |
| | 559 |

US-10-001-857-42 (1-3096) x US-10-168-097A-26 (1-559)

| | | | |
|----|-----|--------------------------------------------------------------|-----|
| QY | 559 | CCATTATGCTA-----CTTCTCCGTGCTACCGCGCGCGCACGCACT | 515 |
| | | ::: ::: | |
| Db | 294 | ProThrCysIleSerSerAlaThrGlyLeuIleGluAsnArgProGlnSerProAlaThr | 313 |
| | | ::: ::: | |
| QY | 514 | GTCGCGACACGGGAANTAACAGCGCGCTCGCGCGCGCGCCCTCCCTCTCAGCC | 455 |

[illegible]

RESULT 7

```

US-10-239-431A-33
; Sequence 33, Application US/10239431A
; Publication No. US20030170726A1
; GENERAL INFORMATION:
; APPLICANT: FRADELIZE, JULIE
; APPLICANT: FRIEDERICH, EVELYNE
; APPLICANT: GOLSTEYN, ROY M.
; APPLICANT: LOUWARD, DANIEL
; APPLICANT: NOIREAUX, VINCENT
; APPLICANT: SYKES, CECILE
; TITLE OF INVENTION: PEPTIDE SEQUENCES COMPRISING ONE OR MORE UNITS BINDING
; TO PROTEINS OF THE Era/VASP FAMILY, AND THEIR USES
; FILE REFERENCE: 0508-1032
; CURRENT APPLICATION NUMBER: US/10/239,431A
; CURRENT FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: PCT/FR01/00843
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: FR 00/03637
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 559
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-239-431A-33

```

| | | |
|------------------------|---------|---------------|
| Alignment Scores: | | |
| Pred. No.: | 0.00942 | Length: |
| Score: | 145.00 | Matches: |
| Percent Similarity: | 40.53% | Conservative: |
| Best Local Similarity: | 31.3% | Mismatches: |
| Query Match: | 2.64% | Indels: |
| DB: | 12 | Gaps: |
| | | 8 |
| | | 559 |

US-10-001-857-42 (1-3096) x US-10-239-431A-33 (1-559)

| | | | | |
|----|-----|--------------------------------------|--------------------------|-----|
| QY | 559 | CAATTATGCCTA-----CTTCTCCCGGGTCA | CCCGCGCGCGACGACCT | 515 |
| | | ::: | | |
| Db | 294 | ProThrCysIleSerSerAlaThrGlyLeuIleGlu | AsnArgProGlnSerProAlaThr | 313 |
| | | ::: | | |
| QY | 514 | GTCGGGACACGGAAATAACAGCGCGCGCGCG | CGCGCGCGCTCCCTCTCAGCC | 455 |
| | | ::: | | |
| Db | 314 | GlyArgThrProValPheValSerThrProPro | ProProProProLeuProSer | 333 |

Qy 454 CAGCCGACGCGCAGCGTGCACGATGATGCCGGGACGCGCGACCCG-----GCC 401
Db |||||
334 AlaLeuSerThrSerSerLeuArgAlaSerMetThrSerProProValPro 353
Qy 400 CCCTCTCCCTTACTCCCGGGACTTCG---CGCTCCCGGGCACACCCCT----- 353
Db |||||
354 ProProProProProProAlaThrAlaLeuGlnAlaProAlaValProProAla 373
Qy 352 -----CCCTCGCCACCG-----CCCTCGGTTC 332
Db |||||
374 ProLeuGlnIleAlaProGlyValLeuHisProAlaProProProIleAlaProProLeu 393
Qy 331 GGCGATGGCGCGCTCTCTTCGCGAGATTAGACGATCGCGAGACCGGAAGTT 272
Db |||||
394 ValGlnProSerProPro-----ValAlaArgAlaAlaProValCys-----GluThr 409
Qy 271 ATCTTTTTCACCTCTGTCGAGTCTGGGTAGTGGCGGTTCGCGAAGAGCGGAG 212
Db |||||
410 ValProValHisProLeuProGlnGlyGluValGlnGlyLeuProProProProPro 429
Qy 211 CCCGAGTCTCAGAGCCCGCCGCTGCTGCGTGCCTCT 173
Db |||||
430 ProProLeu-----ProProGlyIleArgProSer 440

RESULT 8
US-10-116-370-2
; Sequence 2, Application US/10116370
; Publication No. US20030190709A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: PABLO, A POLYPEPTIDE THAT INTERACTS WITH BCL-XL, AND
; FILE REFERENCE: USES RELATED THERETO
; CURRENT APPLICATION NUMBER: US/10116.370
; CURRENT FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 559
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-370-2

Alignment Scores:
Pred. No.: 0.00942 Length: 559
Score: 145.00 Matches: 48
Percent Similarity: 40.52% Conservative: 14
Best Local Similarity: 31.37% Mismatches: 61
Query Match: 2.64% Indels: 30
DB: 12 Gaps: 8

US-10-001-857-42 (1-3096) x US-10-116-370-2 (1-559)

Qy 559 CCATTATGCTTACTTCTCCCGTGTGCACCGCGCGCGCGCGCACT 515
Db |||||
294 ProThrCysIleSerSerAlaThrClyLeuIleGluAsnArgProGlnSerProAlaThr 313
Qy 514 GTCCGGACACGGAATAACGACGCGCTCGCGCGCGCGCGCGCTCCCTCTCAGCC 455
Db |||||
314 GlyArgThrProValPheValSerProThrProProProProProProProSer 333
Qy 454 CAGCCGACGCGCAGCTGCGATGCGTATGCCGGGACGCGGACCC-----GCC 401
Db |||||
334 AlaLeuSerThrSerSerLeuArgAlaSerMetThrSerThrProProProValPro 353
Qy 400 CCCTCTCCCTTACTCCCGGGACTTCG---CGCTCCCGGGCACACCCCT----- 353
Db |||||
354 ProProProProProProAlaThrAlaLeuGlnAlaProAlaValProProProAla 373
Qy 352 -----CCCTCGCCACCG-----CCCTCGGTTC 332
Db |||||
374 ProLeuGlnIleAlaProGlyValLeuHisProAlaProProProIleAlaProProLeu 393

Qy 331 GGCGATGGCGCGCTCTCTTCTCGCGAGATTAGACGATCGCGAGACCGGAAGTT 272
Db |||||
394 ValGlnProSerProPro-----ValAlaArgAlaAlaProValCys-----GluThr 409
Qy 271 ATCTTTTTCACCTCTGTCGAGTCTGGGTAGTGGCGGTTCGCGAAGAGCGGAG 212
Db |||||
410 ValProValHisProLeuProGlnGlyGluValGlnGlyLeuProProProProPro 429
Qy 211 CCCGAGTCTCAGAGCCCGCCGCTGCTGCGTGCCTCT 173
Db |||||
430 ProProLeu-----ProProGlyIleArgProSer 440

RESULT 9
US-10-001-873-50
; Sequence 50, Application US/10001873
; Publication No. US20020160388A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P:
; FILE REFERENCE: DEX-0275
; CURRENT APPLICATION NUMBER: US/10/001.873
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,055
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,496
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 1134
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-873-50

Alignment Scores:
Pred. No.: 0.0145 Length: 1134
Score: 144.00 Matches: 53
Percent Similarity: 37.97% Conservative: 7
Best Local Similarity: 33.54% Mismatches: 64
Query Match: 2.62% Indels: 34
DB: 14 Gaps: 8

US-10-001-857-42 (1-3096) x US-10-001-873-50 (1-1134)

Qy 559 CCATTATGCTTACTTCTCCCGTGTGCACCGCGCGCGCGCACTGTTCGGACACGGA 500
Db |||||
976 ProLeuProLeuArgLeuProProLeuProProProPro----- 988
Qy 499 ATAACGACGCGCTCGCGCGCGCGCGCGCTCCCTCTCA-----GCCAGCCGAC 446
Db |||||
989 ---LeuProArgProHisProProProProProProProProLeuProProLeuProPro 1007
Qy 445 CGGACGCGTGCA-----CGCATGCGTATGCCGGGACGCGGACCGCCGCC 401
Db |||||
1008 GlnThrArgThrLeuProAlaAlaArgThrMetArgGlnPro-----ProPro 1023
Qy 400 CCCTCTCCCTTACTCCCGGGACTTCGCGTCCCGGGCACACCCCTCTCTCGCCA--- 344
Db |||||
1024 ProArgLeuAlaLeuProArgArgSerProProProProProProProProAla 1043
Qy 343 -----CGCTCCCTCGCTCGCGCGCGCGCGCTCTCTCTCTCGCGAGA 299
Db |||||
1044 ArgArgGlyProArgProThrProGlnAlaArgArgProProProProProArg 1063
Qy 298 TTAGAGACGATCGCGAGACCGGAAGTTATCTTTTTCACCTCTGTCTCGGAGTCTCGGTA 239
Db |||||
1064 Leu-----LeuArgSerPro-----HisSerLeuCysSerProArgLeu 1076

| | | | | | |
|-------------------------------------------------------------------|---|--|-----|--------------------------------------------------------------|-----|
| D | b | | 405 | ThrThrIleThrThrThrIle-----ThrThrThrThrIleThrThrThr | 419 |
| Q | y | | 923 | TGGCCTTCAACCAAGTTATCAAACAGCAAAACAATGATCCATCATCCATCAGTTCA | 864 |
| D | b | | 420 | ThrThrThrThrIleThrSerThrThrThrThrThrThrThrThrThrThrThr | 439 |
| Q | y | | 863 | GCGAAGTGAGATCTTAATTATAAGTAGCCATCTTGATAGCTGTTCAAAATTGAGA | 804 |
| D | b | | 440 | -----ThrThrIleThrIleThrSerThrThrThrThrThrThrThrThrThr | 455 |
| Q | y | | 803 | ACTTTTCGATTAACTTGTTTCCCAATCATGCAGCATCCATCTGGGATCCATCTCA | 744 |
| D | b | | 456 | ThrThrThrThrThrIleThrIleThrSerThrThr-----ThrThrThrThr | 472 |
| Q | y | | 743 | ATAGCAGACATGGCTTCAAAAGACCAATAGCTTATCATGAAGTAGTTCTCCCACTT | 684 |
| D | b | | 473 | IleThrThrThrThrThrArgThrThrIleThrThrThrThrThrThrThrThr | 492 |
| Q | y | | 683 | AATTCTGACAAAGTCTCTCAAAATCTTTGGTAATGTCCACCAGTTTGATGTGTTTC | 624 |
| D | b | | 493 | ThrThrIleThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr | 512 |
| Q | y | | 623 | TCCATTITTTGGCATACTAGCTGCCATCCTGAATCGTCAATCTACAGAAGCTTC | 564 |
| D | b | | 513 | ThrThrThrIleThrThrThrThrThrThrThrThrThrThrThrThrThrThr | 532 |
| Q | y | | 563 | ATAACCAATTATGCCTACTTCCCGTGTCAACCGCGCGCACGACACTGTCCGACCA | 504 |
| D | b | | 533 | IleThrIleThrThrThrThrThrIleThrThrThrThrThrThrThrThrThr | 552 |
| Q | y | | 503 | G-----GAAATAAGCAGCCGCGCT-----CG | 483 |
| D | b | | 552 | xThrThrIleThrIleThrSerThrThrThrAlaIleThrThrThrThrThrThr | 572 |
| Q | y | | 502 | -----GAAATAAGCAGCCGCGCT-----CG | 483 |
| D | b | | 572 | eThrIleThrThrThrThrProLeuProProSerProProProProProSerLeuPr | 592 |
| Q | y | | 482 | GCCCGCGCGCGCCCTCCCCCTCTCAGCCACGCCGCGCAGCGTGACCGATCGTAT | 423 |
| u | b | | 592 | oProProProProProSerLeuProProProSerProSerLeuProProProle | 612 |
| Q | y | | 422 | GCCGGGACGCGACCCCGCCCCTCTCCCTTACTC-----CCGGGGAC | 378 |
| D | b | | 612 | uProSerLeuProProProProProProSerProSerLeuProProProProSerPr | 632 |
| Q | y | | 377 | TTGCGCTCCCGGCCACACCCCTCCCTCGTCG---CCACCGCTCGTTCGGCGGATGCG | 321 |
| D | b | | 632 | oserLeuProProProProProProProSerLeuProProProProProSerLeuProPr | 652 |
| Q | y | | 320 | GCGCCCTTCTTCTCGGAGATTAGAGACCATCGGAGACCGGAAGTATCTTTTTTCA | 261 |
| D | b | | 652 | oproProSerLeuProProSerProProSerLeuSerProProSerLeu----- | 668 |
| Q | y | | 260 | CCCTCTCTGGAGTCCTGGGTAGTGGCGGTTCCTCGGAAGCGGACCGCGAGTCTC | 201 |
| D | b | | 669 | -ProLeu-----ProProProProProSerLeuPr | 678 |
| Q | y | | 200 | AGAGCCCGCCCGTCTCGGTGCCC | 176 |
| D | b | | 678 | oproProProProSerLeuProPro | 686 |
| RESULT 11 | | | | | |
| US-10-086-464-5 | | | | | |
| ; Sequence 5, Application US/10086464 | | | | | |
| ; Publication No. US2002019218A1 | | | | | |
| ; GENERAL INFORMATION: | | | | | |
| ; APPLICANT: GORING, Daphne R. et al. | | | | | |
| ; TITLE OF INVENTION: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES | | | | | |
| ; FILE REFERENCE: P 25,762-A USA | | | | | |
| ; CURRENT APPLICATION NUMBER: US/10/086,464 | | | | | |
| ; CURRENT FILING DATE: 2002-02-28 | | | | | |
| ; PRIOR APPLICATION NUMBER: US 10/069,304 | | | | | |

;
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: PCT/CA00/00966
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: US 60/149,466
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: US 60/159,122
; PRIOR FILING DATE: 1999-10-13
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 721
; TYPE: PRT
; ORGANISM: Brassica napus
US-10-086-464-5

Alignment Scores:
Pred. No.: 0.0178 Length: 721
Score: 142.00 Matches: 58
Percent Similarity: 37.04% Conservatives: 22
Best Local Similarity: 26.85% Mismatches: 75
Query Match: 2.58% Indels: 61
DB: 14 Gaps: 10

US-10-001-857-42 (1-3096) x US-10-086-464-5 (1-721)

```
QY 788 TGGTTTCCAATCATCCAGCATCCATCTTGGATCCATCATTTCAATAGCAGACATGGCT 729
Db ||| ||| :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 TrpSerProCysLeuLeuSerProAlaAlaSerPheHisSerLeu----- 19

QY 728 TCAAAAGACCAAAAGCTTATCATGAGTAGTCTTCCCACTTTTAATTCGCAAGCT 669
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20 -----HisLeuProPheArgSerLysGluAla 29

QY 668 TCTTCAAAATCTUGGGTAATGCCACCAGTTTGTATTGCTTTTCTCCATTTTTCGCG 609
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30 AspMetSerSer-----AlaProSerProGlyThrGly 40

QY 608 ATACTGAGCTCCATCTGATCATCATCATCTACA-----GAAGCTTTCATAACCAT 555
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 SerProProSerProProSerAsnSerThrThrThrThrProProProAlaSerAlaPro 60

QY 554 ATGCTACTTCTCCGCTGTACCCGCGCCGCGCAGCTGTCGCGAC----- 506
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 ProProThrThrProSer-SerProProProProSerThrThrThrThrProProPr 80

QY 505 -ACGAAATAAGACGCGCCCTCGCGCGCGCGCCCTCCCTCTCAGCCGCGCGCA 447
Db :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 80 oSerSerArgSerThrProSerAlaProProProSerProProThrProSerThrProGl 100

QY 446 CGGCGAGCTGCACGCATCGTATCCCGGAGCGCGACCGCGC-----CCCTCTCCCT 390
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100 ySerPro-----ProProLeuProGlnProSerProProAlaProTh 114

QY 389 ACTCCCGGGGACTTCGCGTCCCGCGCACACCCCTCCCTCGCCACCGCCTCCGTC-- 332
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114 rThrProGlySerProProAlaProValThrProProThrArgAsnProProProSerVa 134

QY 331 ---GGCGAGTGGCGCGCTTCTCTTCCGAGATAGAGACGATCGGAGACCGGA 276
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134 lProGly-----ProProSer----- 139

QY 275 AGTTATCTTTTACACCTCTGTCGAGTCCTGGAGTCCTGGTGTGGCGGTTCCTCGGAAGAGC 216
Db : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140 -----AsnProSerArgGluGly-----GlySerProArgProProSe 152

QY 215 GGAGCCCGGAGTCTCAGAGCCCGCCCTGCTGCGTGGCGCTCTGGG 170
Db : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 rSerProSerProProSerProSerSerAspGlyLeuSerThrGly 167
```

RESULT 12

US-10-142-515-11
; Sequence 11, Application US/10142515
; Publication No. US20030078399A1

;
; GENERAL INFORMATION:
; APPLICANT: SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH
; APPLICANT: Lloyd, Kenneth O.
; APPLICANT: Yin, Beatrice W.T.
; TITLE OF INVENTION: Nucleic Acid Sequence Encoding Ovarian Antigen, CA125, and Use;
; FILE REFERENCE: 649-A-US
; CURRENT APPLICATION NUMBER: US/10/142,515
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/290,480
; PRIOR FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 5877
; TYPE: PRT
; ORGANISM: Human Being
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(5877)
; OTHER INFORMATION: Amino acid sequence of MUC16B
US-10-142-515-11

Alignment Scores:
Pred. No.: 0.0404 Length: 5877
Score: 141.50 Matches: 148
Percent Similarity: 31.86% Conservatives: 76
Best Local Similarity: 21.05% Mismatches: 266
Query Match: 2.57% Indels: 213
DB: 15 Gaps: 27

US-10-001-857-42 (1-3096) x US-10-142-515-11 (1-5877)

```
QY 2237 CTTCCATTCTCTTCCCTCTGCCATTTCAGAGCCATCGGACGACTCAATGTTGACATCA 2178
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 713 LeuThrLeuAspThrSerThrThrPheMetSerGlyThrHis-----Ser 727

QY 2177 ACCATGCGTAAGGAATTCAGAGATACCAATATATGTAATAGTACTCGTCATACTGT 2118
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 728 ThrAlaSerGlnGlyPheSerHisSerGlnMetThrAlaLeuMetSerArgThrProGly 747

QY 2117 AGAGTTCCAATTCAAAGCCACTTAGAAGGTACTGTATCATATGCGAAGTTATGGTAA 2058
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 748 GluValProTrpLeuSerHisProSer-----ValGluGluAlaSerSer 762

QY 2057 GGACCCAGGTACCTAAACAGCGCAAATGTTCGCTTTGGGGTTCTCTGTTTCAACAGCATGG 1998
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 763 AlaSerPheSerLeuSerSerProValMetThrSerSerSerProValSerThrLeu 782

QY 1997 TGTGAAGCGTGCATCAACCTTCTCTGCTTATTAATAAATGTCATAAATCATCTCCTG 1938
Db :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 783 ProAspSerIleHisSerSerLeuProVal----- 793

QY 1937 CAAGTGGCAATTCCTCAAGATATATACCAAGCTTATCTCTCTCGAGCCCTGTATG 1878
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 794 -----ThrSerLeuLeuThr-----SerGlyLeuValLys 803

QY 1877 TCCATGATCTGAATPAAGACTACAGATGGCCGACACAGTGAACAAAGAGAGTCCGAT 1818
Db :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 804 ThrThrGluLeuLeuGlyThrSerSerGluProGluThrSerSerProProAsnLeuSer 823

QY 1817 ACAGTCTTACGCTGCTATTTATATATAGTAGCACTTGGGGGAAAGCACTCGGAGGAT 1758
Db :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 824 SerThrSerAlaGluIleLeuAlaThrThrGluValThrThrAspThrThrGluLysLeuGlu 843

QY 1757 CTGCAAAAGACCGGAAGTGCATTTTCACCATGTTCTTCATGAGATGAGTTCCTCAAGACC 1698
Db :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 844 MetThrAsnValThrSerGlyThrHisGluSer-----ProSerSer 859

QY 1697 TTTTGTATCC-----ACCAGGAAGTGGTTTGTAAACAGAGATCTTTGAAGAAGACAT 1644
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 860 ValLeuAlaAspSerValThrThrLysAlaThrSerSerMetGlyIleThrThrProThr 879

QY 1643 GGTGACTGTTCACTAAATCA-----CAGAAAAAATCCAGGATACAAATGT 1599
```

```
Db      880 GlyAspThrAsnValLeuThrSerThrProAlaPheSerAspThrSerArgIleGlnThr 899
      ||||| : : : : : |||||
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      1598 AAATTGTTAAATTCACAACTCAGACAGAGTTTATCTATCTATTAACTTCGCAAAA 1539
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      900 Lys----- 900
      ||||| : : : : : |||||
QY      1538 TAGTTACCAATTTCTCCCTTTTAAATTATTTTGCATATCGAGGGAAGGTAGTGAAGT 1479
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      901 -----Ser 901
      ||||| : : : : : |||||
QY      1478 AGCCTCTGGTTCACAAGGGTTCAAAACCCATCATAATTTGGATGATCTCTTTGTAGTA 1419
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      902 LysLeuSerLeuThrProGly----- 908
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      1418 TCATTCTGGGCTCGATGCCATGATGCAATGAATTTATGAATGCGAGAAAGAGATCTGCT 1359
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      909 -----LeuMetGluThrSerIleSerGluGluThrSerSer 920
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      1358 GCTTGAACCATCAATTTTGGAGCTTCTGCAACAGCACTGGTCTCTTTCTTAGTAAAGGCT 1299
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      921 Ala-----ThrGluLysSerThrValLeuSerSerValProThrGlyAla 935
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      1298 ATAAGCACTGTCAGTAACACAGGAGTAATTTCACTCTGCTGAATACATGCTAAACATGTT 1239
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      936 ThrGluValSerArgThrGlu----- 943
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      1238 TGGTGTCTAGTCAACACTCTGGATCTCTTCTCTCTCTGCTTCGACTTCGAGTACTCTTT 1179
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      944 ---AlaIleSerSerArgThrSerIleProGlyProAlaGln----- 957
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      1178 ACTCTCTCTTCATGTCACTCCATCTCT-----TTTAGCATGCT 1137
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      958 -----SerThrMetSerSerAspThrSerMetGluThrIleThrArgIleSerThrPro 975
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      1136 GTRACTCGAAGACTGTCACACTGTAGCCATTTTAATCCATAGTCATTCAGTCAAAA 1077
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      976 LeuThrArgLysGluSerThrAspMetAlaIle---ThrPro-----Lys 989
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      1076 TCTTCTCTCTCAAAAACAGACAGCTTTATTTACTTTTTCCTTCCATGCAATGTCACAGATTTC 1017
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      990 ThrGlyProSerGlyAlaThrSerGlnGlyThrPheThrLeuAspSerSerSerThrAla 1009
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      1016 AAGATTCCAGACAAAGCCCTTCATACAGAGATCTTCTATPAAAGCTTCGGATTATGAATG 957
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1010 SerTrpPro-----GlyThrHisSerAlaThrThrGlnArgPhe 1022
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      956 TAAAGGCAGCAATAACTCTGTCTGCCAGT-----GAATGCCCTCT 915
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1023 ProArgSerValValThrThrProMetSerArgGlyProGluAspValSerTrpProSer 1042
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      914 AAC---CACGTTATCAACACAGCAAAAACATGTATCCATATCCCTATCAGTTTCAGGCAAG 858
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1043 ProLeuSerValGluLysAsnSerProProSerSerLeuValSerSerSerValThr 1062
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      857 GTGAGATCTTTAATTTAATAGTCCATCTCTGTAGTACTTGTTCAAAATTGAGAACTTTT 798
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1063 SerProSerProLeuTySerThrProSerGlySerSerHisSerSer----- 1078
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      797 CGATTAACTTGTTTCCAAATCATCCAGCATCCATCTTGGATCCATCATCTTCAATAGCA 738
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1079 -----ProValProValThrSerLeuPheThrSerIleMetMetLysAla 1093
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      737 ---GACATG-----GCTTCAAAAAGACCAATAGCTTATCATGAAGTAGTTCCTCCCAAC 687
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1094 ThrAspMetLeuAspAlaSerLeuGluProGluThr-----ThrSerAlaProAsn 1110
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      686 TTTAATTCTGCAGAACTCTTCAAAATCTGGTAATGTCACCACTGTTGTATTGCTT 627
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1111 MetAsnIleThrSerAspGluSerLeuAlaAlaSerLysAlaThrThrGluThrGluAla 1130
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      626 TCTCCATTTTCTTCTGCACTACTGAGCTCCCATCTCTGAATCGTCATCATCTACGAAGCT 567
      ||||| : : : : : |||||
      ||||| : : : : : |||||
```

```
Db      1131 IleHisValPheGluAsnThrAlaAlaSerHisValGluThrThrSerAlaThrGluGlu 1150
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      566 TTCATAACCATATATGCTT-----ACTTCTCCCTGGTGCACC 531
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1151 LeuTySerSerSerProGlyPheSerGluProThrLysValIleSerProValValThr 1170
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      530 CGGCGCGCCACGCACTGTCTCGGACCAAGAAATACAGCGCGCTCGCGCCGCCGCCGCC 471
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1171 -----SerSerSerIleArg 1175
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      470 CCCTCCCTCTCAGCCAGCCCGACCGGAGCGTGTCAGCGCATGCTATGCCCGGACGCG 411
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1176 AspAsnMetValSerThrThrMetProGly-----SerSerGlyLeuThrArgIleGlu 1193
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      410 CGACCCCGCCCTCTCTCTTACTCCCGGAGACTTCCTCGGCTCCCGGCGCACACCCCTCC 351
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1194 IleGluSerMetSerSerLeuThrProGly----- 1203
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      350 CTCGCGCACCGCTCGTTCGCGCGCATGCGCGCGCTTCTCTTCGCGAGATTAGAGAC 291
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1204 -----LeuArgGluThrArgThrSerGlnAspIleThrSer 1215
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      290 GATCGGAGACCGGAAGTTATCTTTTTCACCCCTCTGTCTCGGAGTCTCTGGTAGTGGCGG 231
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1216 SerThrGluThrSerThrValLeuTyLys-----MetProSerGlyAla 1230
      ||||| : : : : : |||||
      ||||| : : : : : |||||
QY      230 TTCGCGGAA 222
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      1231 ThrProGlu 1233

RESULT 13
US-10-243-243A-8
; Sequence 8, Application US/10243243A
; Publication No. US2003010442A1
; GENERAL INFORMATION:
; APPLICANT: Lloyd, Kenneth O.
; TITLE OF INVENTION: Nucleic Acid Sequence Encoding Ovarian Antigen, CA125, and Use.
; FILE REFERENCE: 649-B
; CURRENT APPLICATION NUMBER: US/10/243,243A
; CURRENT FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: US 10/142,515
; PRIOR FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: PCT/US02/14768
; PRIOR FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: US 60/290,480
; PRIOR FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 5935
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)-(5935)
; OTHER INFORMATION: Amino acid sequence of MUC16B
US-10-243-243A-8

Alignment Scores:
Pred. No.: 0.0406 Length: 5935
Score: 141.50 Matches: 148
Percent Similarity: 31.86% Conservative: 76
Best Local Similarity: 21.05% Mismatches: 266
Query Match: 2.57% Indels: 213
DB: 15 Gaps: 27

US-10-001-857-42 (1-3096) x US-10-243-243A-8 (1-5935)

QY      2237 CTTCATTATCTTCTCTCGCCATTGAGAGCATCGGACGACTCAATGTTGACATCA 2178
      ||||| : : : : : |||||
      ||||| : : : : : |||||
Db      784 LeuThrLeuAspThrSerThrThrPheMetSerGlyThrHis-----Ser 798
      ||||| : : : : : |||||
      ||||| : : : : : |||||
```

2177 ACCATGCGTAAAGAAITTCAGAGAGATACCAATATATAGTACTCGTGCACTGT 2118
Db
799 ThrAlaSerGlnGlyPheSerHisSerGlnMetThrAlaLeuMetSerArgThrProGly 818
Qy
2117 AGAGTTCAATTCAAAGCACTTAGAGGTACTGATCATATGCGAAGGTATGGTAA 2058
Db
819 GluValProTrpLeuSerHisProSer-----ValGluGluAlaSerSer 833
Qy
2057 GGACCCAGGTACCTAAACAGCGCAAAATGTCCTTTGGGGTTCTGTTTCAACAGCATGG 1998
Db
834 AlaSerPheSerLeuSerSerProValMetThrSerSerSerProValSerSerThrLeu 853
Qy
1997 TGTGAAGCGTGCATCAACCTTCTGCTGCTATTAAATPAAATGTCATAAACTCATCTG 1938
Db
854 ProAspSerIleHisSerSerLeuProVal----- 864
Qy
1937 CAAGTGGCAATTCCTCAAGAAATGACCAAGCTTATCTCTGTCGAGCCCTGTTATG 1878
Db
865 -----ThrSerLeuLeuThr-----SerGlyLeuValLys 874
Qy
1877 TCCATGGATCTGAATAAGACTACAGAATGGCGCAACACAGTGAATAACAAAGGAGTCGAT 1818
Db
875 ThrThrGluLeuLeuGlyThrSerSerGluProGluThrSerSerProAsnLeuSer 894
Qy
1817 ACAGTCTTAGCTGCTGATTAATATATAGTAGCACTTGGGGAAAGCACTCGGAGAT 1758
Db
895 SerThrSerAlaGluIleLeuAlaThrThrGluValThrThrAspThrGluLysLeuGlu 914
Qy
1757 CTGCAAAAGACCGAAGTGCATCTTTCACCATGTCTTGCAATGAGATGAGTCCAAAGACC 1698
Db
915 MetThrAsnValThrSerGlyThrHisGluSer-----ProSerSer 930
Qy
1697 TTTTGTGTTATCC-----ACGAGAAAGTGGTTTGTAAACAGAGATCTTGAAGAACACAT 1644
Db
931 ValLeuAlaAspSerValThrThrLysAlaThrSerSerMetGlyIleThrTyrrProThr 950
Qy
1643 GGTGACTGTTCTCAATTCAC-----CAGAAAAATCCAGGATACAAATGT 1599
Db
951 GlyAspThrAsnValLeuThrSerThrProAlaPheSerAspThrSerArgIleGlnThr 970
Qy
1598 AAATTTGTTAAATTCACACCTTCACAGACAGTGTATTTATCTATCTATTATCTGCAAAA 1539
Db
971 Lys----- 971
Qy
1538 TAGTTCAACCATTTCTTCCCTTTTAATTAATTTTGCATATCGAGGAAGTAGTGGAAGT 1479
Db
972 -----Ser 972
Qy
1478 AGCTTCTGGTTCAAGGGGTTCAAAACCCATCAATAATGGGATGATCTCTTTGTAGTA 1419
Db
973 LysLeuSerLeuThrProGly----- 979
Qy
1418 TCATTCTGGCGCTGGATGCCATGATGCAATGAATATGAATGCGAGAAAGATGCTGT 1359
Db
980 -----LeuMetGluThrSerIleSerGluGluThrSerSer 991
Qy
1358 GCTTGAACCATCAATTTTGGAGCTTCTGCAACAGCACTGCTCTCTTTCTTGTAAAGGCT 1299
Db
992 Ala-----ThrGluLysSerThrValLeuSerSerValProThrGlyAla 1006
Qy
1298 ATAAGCACTGTCAGTAACACAGCAGTAAATTTTCACTCTCTGTAATCTGCTAAACATTGT 1239
Db
1007 ThrThrGluValSerArgThrGlu----- 1014
Qy
1238 TGGTGTCTAGTTCAACTTCTGGAATCTCTTCTCTGCTGTCGACTCGAGTACTCTTT 1179
Db
1015 ---AlaIleSerSerSerArgThrSerIleProGlyProAlaGln----- 1028
Qy
1178 ACTCTTCTTGTGATGTCATCTCCACATCT-----TTTAGCATGCT 1137
Db
1029 -----SerThrMetSerSerAspThrSerMetGluThrIleThrArgIleSerThrPro 1046
Qy
1136 GTAACCTCGAAGATCTGTCACACTGTTAGCATTGTTAAATCCATTAAGTCATTGACTGAAAA 1077

Db
1047 LeuThrArgLysGluSerThrAspMetAlaIle-----ThrPro-----Lys 1060
Qy
1076 TCTTCTCTTCAAAAACAGCAGCTTTATTTACTTTTCCCTTGCATATGTCACAGATTTTC 1017
Db
1061 ThrGlyProSerGlyAlaThrSerGlnGlyThrPheThrLeuAspSerSerSerThrAla 1080
Qy
1016 AAGATTCACAGCAAAAGCCTTCATGACAGATCTTCTATAAAAGTCTGGATATGATG 957
Db
1081 SerTrpPro-----GlyThrHisSerAlaThrThrGluArgPhe 1093
Qy
956 TAAAGGACGCAATAATACTGTCTGTGCCAGT-----GAATGGCTTCT 915
Db
1094 ProArgSerValValThrThrProMetSerArgGlyProGluAspValSerTrpProSer 1113
Qy
914 AAC---CACGTTTATCAACACAGCAAAACATGTATCCATAATCCCTATCAGTTCAGGCAAG 858
Db
1114 ProLeuSerValGluLysAsnSerProProSerSerLeuValSerSerSerValThr 1133
Qy
857 GTGAGATCTTTAATTTTAAATAGTCCATCTTGTAGTCTGTTCAAAATGAGAACTTTT 798
Db
1134 SerProSerProLeuTyrrSerThrProSerGlySerHisSer----- 1149
Qy
797 CGATTAACTTGGTTTCCGAATCATGCCAGCATCCATCTTGGGATCCATCATTTCAATAGCA 738
Db
1150 -----ProValProValThrSerLeuPheThrSerIleMetMetLysAla 1164
Qy
737 ---GACATG-----GCTTCAAAAAGACCAATAGCTTATCATGAGTAGTTCCTCCCAAC 687
Db
1165 ThrAspMetLeuAspAlaSerLeuGluProGluThr-----ThrSerAlaProAsn 1181
Qy
686 TTTAATCTCGCAAGCTTCTTCAAAATCTTGGGTAATGTCCACCAGTGTGATGCTT 627
Db
1182 MetAsnIleThrSerAspGluSerLeuAlaAlaSerLysAlaThrThrGluThrGluAla 1201
Qy
626 TTCCTCATTTTCTGGCATACTGAGTCCCATCTCTGAATCTGTCATCATCTACAGAAGCT 567
Db
1202 IleHisValPheGluAsnThrAlaAlaSerHisValGluThrThrSerAlaThrGluGlu 1221
Qy
566 TTCATAACCATTTATGCT-----ACTTCTCCCGTGGTCAAC 531
Db
1222 LeuTyrrSerSerSerProGlyPheSerGluProThrLysValIleSerProValThr 1241
Qy
530 CGCGCCGCCACGCACTGTCCGAGACACCGAAATAACGACGCGCTCGCGCCGCCCGCC 471
Db
1242 -----SerSerIleArg 1246
Qy
470 CCTTCCCTCTCAGCCCGACCGCGGCGTGCACGATGCGTATGCCCGGACGCG 411
Db
1247 AspAsnMetValSerThrThrMetProGly-----SerSerGlyIleThrArgIleGlu 1264
Qy
410 CGACCCCGCCCTCTCCCTTACTCCCGGGGACTTTCGCGCTCCCGGCCACACCCCTCC 351
Db
1265 IleGluSerMetSerSerLeuThrProGly----- 1274
Qy
350 CTCGCCACGCTCCGTCGCGCGGATGGCGCGCTTCTCTTCTCGCGAGATTAGAGAC 291
Db
1275 -----LeuArgGluThrArgThrSerGlnAspIleThrSer 1286
Qy
290 GATCGCGAGACCGAAGTTATCTTTTTCACCTCTGTGCGAGTCTCGGTAGTGGCGG 231
Db
1287 SerThrGluThrSerThrValLeuTyrrLys-----MetProSerGlyAla 1301
Qy
230 TTCGCGAA 222
Db
1302 ThrProGlu 1304

RESULT 14

US-09-823-240-2
; Sequence 2, Application US/09823240
; Patent No. US2002004813A1
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gertler


```
; APPLICANT: James E. Bear
; APPLICANT: Jurgen Wehland
; APPLICANT: Joseph Loureiro
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; TITLE OF INVENTION: Motility
; FILE REFERENCE: M0656/7064 (HCL)
; CURRENT APPLICATION NUMBER: US/09/823,240
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,564
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-823-240-2
Alignment Scores:
Pred. No.: 0.0267 Length: 802
Score: 140.00 Matches: 49
Percent Similarity: 36.24% Conservative: 5
Best Local Similarity: 32.89% Mismatches: 58
Query Match: 2.55% Indels: 37
DB: 9 Gaps: 6
US-10-001-857-42 (1-3096) x US-09-823-240-2 (1-802)
QY 535 TCACCCGCGCGGCACGCACACTGTCGGGACACGAAATACGACGCGCCCTCGCGCGCG 476
Db 532 SerProThrProGlnGlyLeuVal-----LeuGlyProProAlaProPro 546
QY 475 CCGCCCCCTCTCTCAGCCCGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 416
Db 547 ProProProProLeuProSerGlyProAlaTyAlaSerAlaLeuProProProGly 566
QY 415 ACGGCGACCGCGCCCTCTCTCTTACTCCCGGGACTTCGCGCTCCCGCGGCACACCC 356
Db 567 -----ProProProProProProProProLeuProSerThrGly-----ProProPro 580
QY 355 CCTCCCTCGCACCGCTCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 296
Db 581 ProProProProProProProProLeuProAsnGlnAlaProProPro----- 595
QY 295 GAGACGATCGGACGCGGAGGTTATCTTTTTCACCCCTCTGTCGGAGTCTCGGTTAGTG 236
Db 596 -----ProProProProProProProProAlaProProLeuProAlaSerGlyIlePhe 610
QY 235 GCGGTTCCCGGAAGAG-----GCGGAGCCCGGA 206
Db 611 SerGlySerThrSerGluAspAsnArgProLeuThrGlyLeuAlaAlaIleAlaGly 630
QY 205 GTCTCAGAGCCCGCCCGCTCGGCTCGCCCTCTGGGAATTTCTAGAAAGGTGACGCGAA 146
Db 631 AlAlaLeuArgLysValSerArgValGluAspGlySerPhe-----Pro 645
QY 145 GCGGAGGAAGAACCGTGGAGTTTG 121
Db 646 GlyGly-GlyAsnThrGlySerVal 653
RESULT 15
US-10-196-935A-4
; Sequence 4, Application US/10196935A
; Publication No. US20030082720A1
; GENERAL INFORMATION:
; APPLICANT: Lifton, Richard P
; APPLICANT: Wilson, Frederick H
; APPLICANT: Choate, Keith
; APPLICANT: Ishikawa, Kazuhiko
; APPLICANT: Nelson-Williams, Carole
; TITLE OF INVENTION: COMPOSITIONS METHODS AND KITS RELATING TO TREATING AND DIAGNOSING
; TITLE OF INVENTION: HYPERTENSION
; FILE REFERENCE: 044574-5113
```

```
; CURRENT APPLICATION NUMBER: US/10/196,935A
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: US 60/306,084
; PRIOR FILING DATE: 2001-07-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-196-935A-4
Alignment Scores:
Pred. No.: 0.0341 Length: 1243
Score: 139.50 Matches: 61
Percent Similarity: 35.24% Conservative: 13
Best Local Similarity: 29.05% Mismatches: 63
Query Match: 2.54% Indels: 73
DB: 15 Gaps: 10
US-10-001-857-42 (1-3096) x US-10-196-935A-4 (1-1243)
QY 535 TCACCCGCGCGGCACGCACACTGTC-----CGGACACGGAATAACGACGCGCCCTCGG 482
Db 4 SerProAlaThrGluThrValLeuMetSerGlnThrGluAlaAspLeuAlaLeuArg 23
QY 481 CCGCGCGCGCGCC-----CCTCCCTCTCAGCCCGCCGCGCGCGCGCGAG 440
Db 24 ProProProProLeuGlyThrAlaGlyGlnProArgLeuGlyProProProArgAla 43
QY 439 CGT-----GCACGCATCGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 410
Db 44 ArgArgPheSerGlyLysAlaGluProArgProArgSerSerArgLeuSerArgSer 63
QY 409 -----GACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 392
Db 64 SerValAspLeuGlyLeuLeuSerSerTipSerLeuProAlaSerProAlaProAspPro 83
QY 391 -----TTACTCCCGGGACTTCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 350
Db 84 ProAspProProAspSerAlaGlyProGlyProGlyProGlyProGlyProGlyPro 101
QY 349 TCGCCACCGCGCTCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 290
Db 102 SerLysGluProProGluGlyThrTrpThrGluGlyAlaProValLysAlaAlaGluAsp 121
QY 289 ATCGCGAGACCGGAAGTTATCTTTTTCACCCCTCTGTCGGAGTCTCGGTTAGTGCGGT 230
Db 122 SerAlaArgProGlu-----LeuProAspSerAlaValGlyProGly 135
QY 229 TCCCGGAAAGCGCGGAGCGCGGAGTCTCAGAGCCCGCGCGCGCGCGCGCGCGCGCG 170
Db 136 SerArgGlu-----ProLeuArgValProGluAla 145
QY 169 AATTTCTTAGAAAGGTGACGCGAAGCGGAAGAACCGTGGAGTTTTCGTAGACCTTA 110
Db 146 ValAlaLeuGluArgArgArgGlu----- 153
QY 109 GTGGTTGCGAGCACGAACGAGCAAGCTCCGAGAGCAACAGCGTCTACTGCTGCTCT 50
Db 154 -----GlnGluGluLysGluAspMetGluThrGlnAlaValAlaThrSerPro 169
QY 49 GACACCGATTATG-----TTTCCTATACAG 23
Db 170 AspGlyArgTyrLeuLysPheAspIleGlu 179
Search completed: November 25, 2003, 03:26:56
Job time : 155 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2003, 21:46:55 ; Search time 6196 Seconds

(without alignments)
12144.393 Million cell updates/sec

Title: US-10-001-857-42

Perfect score: 3096

Sequence: 1 ttcttcacgaactccagg.....atttcgtgacaaaaaaa 3096

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_htc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: em_gss_hum:**

18: em_gss_inv:**

19: em_gss_pln:**

20: em_gss_vit:**

21: em_gss_fun:**

22: em_gss_man:**

23: em_gss_mus:**

24: em_gss_pro:**

25: em_gss_rod:**

26: em_gss_pbg:**

27: em_gss_vrl:**

28: gb_gss1:**

29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 795 | 25.7 | 1033 | 11 | BC027201 Mus muscu |
| 2 | 777.8 | 25.1 | 973 | 11 | AK021042 Mus muscu |
| 3 | 705 | 22.8 | 722 | 12 | BM979022 UI-CF-DUI |
| 4 | 696.8 | 22.5 | 834 | 10 | BF969365 602271630 |

| | | | | | | |
|---|----|-------|------|------|----|----------|
| C | 5 | 658.6 | 21.3 | 814 | 14 | CB169246 |
| | 6 | 653.8 | 21.1 | 3071 | 11 | AK050805 |
| | 7 | 653.6 | 21.1 | 768 | 10 | BG547889 |
| | 8 | 648 | 20.9 | 957 | 13 | BQ965085 |
| C | 9 | 641.4 | 20.7 | 828 | 9 | AI790514 |
| | 10 | 610 | 19.7 | 616 | 12 | BM979511 |
| C | 11 | 599.2 | 19.4 | 900 | 14 | CA977646 |
| | 12 | 598.8 | 19.3 | 697 | 12 | BI602827 |
| C | 13 | 597 | 19.3 | 598 | 9 | AM007583 |
| | 14 | 593.8 | 19.2 | 718 | 13 | BU058931 |
| | 15 | 586 | 18.9 | 972 | 10 | BG112812 |
| | 16 | 584.8 | 18.9 | 2958 | 11 | BC043701 |
| | 17 | 575.2 | 18.6 | 763 | 13 | BU613022 |
| | 18 | 572.4 | 18.5 | 952 | 13 | BQ925670 |
| C | 19 | 560 | 18.1 | 560 | 9 | AA888870 |
| | 20 | 555.6 | 17.9 | 927 | 11 | BC010475 |
| | 21 | 551.2 | 17.8 | 1117 | 13 | BU525571 |
| | 22 | 548.4 | 17.7 | 873 | 9 | AU081000 |
| | 23 | 542 | 17.5 | 735 | 13 | BU057675 |
| | 24 | 538.6 | 17.4 | 654 | 10 | BF468052 |
| | 25 | 537 | 17.3 | 537 | 12 | BM699453 |
| | 26 | 526.8 | 17.0 | 666 | 13 | BQ443173 |
| | 27 | 523.2 | 16.9 | 637 | 14 | CD372119 |
| | 28 | 507.4 | 16.4 | 509 | 10 | BG612374 |
| | 29 | 506.8 | 16.4 | 906 | 12 | BI658829 |
| | 30 | 503.2 | 16.3 | 1114 | 10 | BG245408 |
| | 31 | 499.6 | 16.1 | 640 | 13 | BQ443252 |
| | 32 | 496.6 | 16.0 | 1049 | 13 | BQ885743 |
| | 33 | 495.6 | 16.0 | 631 | 14 | CB418817 |
| | 34 | 491.4 | 15.9 | 690 | 10 | AW975643 |
| | 35 | 490.8 | 15.9 | 901 | 12 | BI731865 |
| | 36 | 487 | 15.7 | 494 | 12 | BM510376 |
| | 37 | 482.2 | 15.6 | 721 | 9 | AJ453000 |
| | 38 | 480.2 | 15.5 | 615 | 14 | CA749517 |
| C | 39 | 472.6 | 15.3 | 479 | 9 | AI339059 |
| C | 40 | 470.4 | 15.2 | 485 | 9 | AW090031 |
| C | 41 | 465 | 15.0 | 480 | 14 | CB242727 |
| | 42 | 464.8 | 15.0 | 556 | 10 | BG062635 |
| | 43 | 463.2 | 15.0 | 558 | 14 | CA555718 |
| | 44 | 461.6 | 14.9 | 569 | 12 | BI682349 |
| | 45 | 461.4 | 14.9 | 784 | 13 | BU260643 |

ALIGNMENTS

RESULT 1

BC027201

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC027201 1033 bp mRNA linear HTC 20-SEP-2002
Mus musculus, Similar to corneal wound healing related protein,
clone IMAGE:483189, mRNA.

BC027201 GI:20070883

HTC.

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1033)

Strausberg,R.

Direct Submission

Submitted (04-APR-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgaabs@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

QY 1381 CATAATTCATTCATCGCATCGCATCGCAGCCAGCAATGATGATACAAAGGAGATCATCCA 1440
 Db |||||
 QY 1441 ATTATGATGGGTTTGAACCCCTTTGTAACAGAGGCTACTTCCACCTACCTTCCCTCGA 1500
 Db |||||
 QY 1501 TATGC 1505
 Db |||||
 QY 969 TATGC 973

RESULT 3
 BM979022/c
 LOCUS 602271630F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4359702 5',
 DEFINITION mRNA sequence.
 ACCESSION BF969365
 VERSION BF969365.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 722)
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 8889548
 CONTACT: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul-mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com) or from Open Biosystems
 (www.openbiosystems.com).
 Seq primer: M13 FORWARD
 POLyA=Yes.

FEATURES
 source
 1..722
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-DUI-1-adl-c-13-0-UI"
 /tissue_type="Primary Lung Epithelial Cells"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-CF-DUI"
 /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-CF-DUI is a normalized cDNA library containing the
 following tissue(s): Primary Lung Epithelial Cells The
 Soares, Genome Research, 6:791-806, 1996. First strand
 cDNA synthesis was primed with an oligo-dT primer
 containing a Not I site. Double stranded cDNA was ligated
 to an EcoR I adaptor, digested with Not I, and cloned
 directionally into pT7T3-Pac vector. The oligonucleotide
 used to prime the synthesis of first-strand cDNA contains
 a library tag sequence that is located between the Not I
 site and the (dT)18 tail. The sequence tag for this
 library is GGCTGTAGGC.
 TAG_LIB=UI-CF-DUI

BASE COUNT 215 a 147 c 124 g 236 t
 ORIGIN
 Query Match 22.8%; Score 705; DB 12; Length 722;
 Best Local Similarity 99.9%; Pred. No. 7.2e-117;
 Matches 716; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 2380 CAAAGTACGTAAACCGAAGTTTGAGCTTGATGCAACAAAGTTCGGTATGACACAGGTT 2439
 Db |||||
 QY 722 CAAAGTACGTAAACCGAAGTTTGAGCTTGATGCAACAAAGTTCGGTATGACACAGG-T 664
 QY 2440 TGCTCCATTCAACAGTGTGATGACCCCGCCGACGTGCACTACTTACAGTTCAAGAAAT 2499
 Db |||||
 QY 663 TGCTCCATTCAACAGTGTGATGACCCCGCCGACGTGCACTACTTACAGTTCAAGAAAT 604
 QY 2500 GTCTGACCTCAATAAATATAGCCCTCCCTCAGTCTCCTGAACTGTATGTGGCAGCTAG 2559
 Db |||||
 QY 603 GTCTGACCTCAATAAATATAGCCCTCCCTCAGTCTCCTGAACTGTATGTGGCAGCTAG 544
 QY 2560 TAAGCACCTTTCAACAGGCAAAATGATATGGAATAATTTCTTAACCCGACCATGAGT 2619
 Db |||||
 QY 543 TAAGCACCTTTCAACAGGCAAAATGATATGGAATAATTTCTTAACCCGACCATGAGT 484
 QY 2620 TAATAGAAATTTAAAGGTTGCCAAACCCAACTTTTGTGGTTATGAAGTTATTGGCAGGAG 2679
 Db |||||
 QY 483 TAATAGAAATTTAAAGGTTGCCAAACCCAACTTTTGTGGTTATGAAGTTATTGGCAGGAG 424
 QY 2680 ACACAAAAGGAATCTAAAGTTCTTCTGAAATTTGATTTCTCTGCTCATAAATATTTTCC 2739
 Db |||||
 QY 423 ACACAAAAGGAATCTAAAGTTCTTCTGAAATTTGATTTCTCTGCTCATAAATATTTTCC 364
 QY 2740 TGTGTGAAACTTGTTCAGAGAGACTGGGAGGTCGCATAAAGGGGACAGATCTTCTT 2799
 Db |||||
 QY 363 TGTGTGAAACTTGTTCAGAGAGACTGGGAGGTCGCATAAAGGGGACAGATCTTCTT 304
 QY 2800 TCAGACCCAACTCTTTAGAGGGCACATCACCAGGCTCCACATCACCAGGAGTGGAT 2859
 Db |||||
 QY 303 TCAGACCCAACTCTTTAGAGGGCACATCACCAGGCTCCACATCACCAGGAGTGGAT 244
 QY 2860 TTCTTGGGTAACTCATTTAAGGAATACCTTTAGTTTGGACGCTTATATGACATGA 2919
 Db |||||
 QY 243 TTCTTGGGTAACTCATTTAAGGAATACCTTTAGTTTGGACGCTTATATGACATGA 184
 QY 2920 ATGAAACTGCTGTTTAAAGTGGTTTATTTATGTTCCATGGAAGAACTGGTCTTATTCGA 2979
 Db |||||
 QY 183 ATGAAACTGCTGTTTAAAGTGGTTTATTTATGTTCCATGGAAGAACTGGTCTTATTCGA 124
 QY 2980 ATGCATTGATGAACGTTATATGTTTATTTACAGATTTAATCACAATCATTTTTATGA 3039
 Db |||||
 QY 123 ATGCATTGATGAACGTTATATGTTTATTTACAGATTTAATCACAATCATTTTTATGA 64
 QY 3040 ATGATTGATGAATAGTGTTTAATAGGTTAATAAATTTCTTACAAAAA 3096
 Db |||||
 QY 63 ATGATTGATGAATAGTGTTTAATAAGGTTAATAAATTTCTTACAAAAA 7

RESULT 4
 BF969365
 LOCUS 602271630F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4359702 5',
 DEFINITION mRNA sequence.
 ACCESSION BF969365
 VERSION BF969365.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 854)
 NIH-MGC http://mgc.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE

| | | | |
|------------|--------------------------------------------------------------------|---------------------------------------------------------------------|-----------------|
| QY | 1467 | GAACGAGCGCTACTTCCACCTACCTCCCTCGATATGCAAAAAATAATTAAGGGAAGA | 1526 |
| Db | 694 | TAACCAAGACTACTTCCACCCACCTTCCCTCGATGCAAAAAATAAAGAGAGA | 635 |
| QY | 1527 | AATGTTGAAGTATTTTGCAGGATTAATAGATAGATATAAAACCTGCTGTGAGGTTGTGAA | 1586 |
| Db | 634 | AATGTTCAACTATTTTGCAGGATTTGATAGATAGATATAAAACCTGCTGTGAGGTTGTCAA | 575 |
| QY | 1587 | TTTAAACAAATTTACATTTGATCTCTGGATTTTCTGTGAATTTAGTGAACAGTCCACATG | 1646 |
| Db | 574 | TTTAAACAAATTTACATTTGATCTCTGGATTTTCTGTGAATTTAGTGAACATTCACCTTG | 515 |
| QY | 1647 | TGTTCTTTCAAGATCTCTTTAAACACCTTTCCCTGGTGGATAACAAAAAGCTCTTTGG | 1706 |
| Db | 514 | TGTTCTTTCAAGATCTCTTTAAACACCTTTCCCTGGTGGATAACAAAAAGCTCTTTGG | 455 |
| QY | 1707 | RACTCATCTCATGCAAGCATGGTGAAGATGCATCTCGTCTTTTGTGAGTCTCTCCGA | 1766 |
| Db | 454 | AATCCATCTCATGCAAGCATGGTGAAGATGCTCTTCGGTCTTTTGTGAGTCTCTCCGG | 396 |
| QY | 1767 | GTGCTTTCCCCCAAGTGCTACCTATATATAATATCAACGGCTAAGGACTGTATCGACTCC | 1826 |
| Db | 395 | -TGCTCTCCCAAGTGCTCCCTATATATAATATCAACGGCTAAGGACTGTATGACTCT | 337 |
| QY | 1827 | TTTGTACTACTGTGTTCCGCCATCTCTGTAGTCTTTATTCAGATCCATCGACATACAGG | 1886 |
| Db | 336 | TTTGTACTACTGTGTTCCGCCATCTCTGTAGTCTTTATTCAGATCCATCGACATACAGG | 277 |
| QY | 1887 | GCTCGACAGAGATAAGCTTGTCATATCTTTGAGGAATTTGCCACCTTGCAGGATGAG | 1946 |
| Db | 276 | GCTCGCAGAGATATAACTCGGTATATCTTTGAGGAATTTGCTACCTTGCAGGATG-- | 219 |
| QY | 1947 | TTTATGACATTTTATTTTATAGGCAGAGAAGTTGATCGAGCGTTTCACACCATCTCT | 2006 |
| Db | 218 | -----AGGCAGAGAAGTTGATCGAGCGCTTTCATCTACTATGCTCT | 180 |
| QY | 2007 | TGAACAGGAACCCAAAGCACATTTGCGCTGTTTAGGTACCTGGTCCCTTTACCATTA | 2066 |
| Db | 179 | TGAACAGAGAGCCCCACAGCGACATTTGGCTGTTTAGGTACCTGGTCCCTTTACCATTA | 120 |
| QY | 2067 | ACCTTCGCATTATGATACAGTACCTTCTAAGTGGCTTTGAAATGGAACTCTA-CAGTATG | 2125 |
| Db | 119 | ACCTTCGAATTATGATACAGTATCTCTAAGTGGCTTCGAATGGAGCTCTACCAATG | 60 |
| QY | 2126 | CACGAGTACTATTACATATTTGGTATCTCTCGAATTCCTTTACGCATGGTTGATGTC | 2184 |
| Db | 59 | CACGAGTATTATACATATATTGGTATCTCTCTGAATTTCTTATGCTGCTGGTTGATGTC | 1 |
| RESULT 6 | | | |
| AK050805 | | | |
| LOCUS | | | |
| DEFINITION | AK050805 | 3071 bp mRNA linear | HTC 05-DEC-2002 |
| | | Mus musculus 9 days embryo whole body cDNA, RIKEN full-length | |
| | | enriched library, clone:D030020M24 product:CORNEAL WOUND HEALING | |
| | | RELATED PROTEIN homolog [Rattus norvegicus], full insert sequence. | |
| ACCESSION | AK050805 | | |
| VERSION | AK050805.1 | GI:26094130 | |
| KEYWORDS | HTC; CAP trapper. | | |
| SOURCE | Mus musculus (house mouse) | | |
| ORGANISM | Mus musculus | | |
| | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | |
| | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| REFERENCE | | | |
| AUTHORS | 1 | Carninci, P. and Hayashizaki, Y. | |
| TITLE | | High-efficiency full-length cDNA cloning | |
| JOURNAL | | Meth. Enzymol. 303, 19-44 (1999) | |
| MEDLINE | | 99279253 | |
| PUBMED | | 10349636 | |
| REFERENCE | | 2 | |
| AUTHORS | | Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., | |
| | | Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. | |
| TITLE | | Normalization and subtraction of cap-trapper-selected cDNAs to | |
| | | prepare full-length cDNA libraries for rapid discovery of new genes | |

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, Y.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S.
and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
5

The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6 (bases 1 to 3071)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, Y., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saichoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsukumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.

| | |
|---------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/. | |
| FEATURES | Location/Qualifiers |
| source | 1..3071 |
| | /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM_DB:D030020M24" /db_xref="taxon:10090" /clone="D030020M24" /tissue_type="whole body" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="9 days embryo" |
| misc_feature | 1..3071 |
| | [note="CORNEAL WOUND HEALING RELATED PROTEIN homolog (Rattus norvegicus) (SPTR Q9J101, evidence: FASTY, 96.2%ID, 43.5%length, match=948)"] |
| BASE COUNT | 811 a 611 c 686 g 963 t |
| ORIGIN | |
| | Query Match 21.1%; Score 653.8; DB 11; Length 3071; Best Local Similarity 88.1%; Pred. No. 1e-107; Matches 736; Conservative 0; Mismatches 92; Indels 7; Gaps 2; |
| Qy | 1316 AGACGAGTGTGTCGAGAGCTCAAAATGATGGTTCAAGCAGCAGATCTTCTTCTG 1375 |
| Db | 210 AGACGAGTGTGTCGAGAGCTCAGAACTGATGGTCCAGCGGCAGACCTTCTTCTG 269 |
| Qy | 1376 CCATTCAATTCATTCGATCATCGCATCCAGGCCCGCAGATGATCACTACAAAGGAGATC 1435 |
| Db | 270 CCATTCACTCATTCGACCCAGGCATCCAGGCTCAGATGGCAGTACCAGAGGAGACC 329 |
| Qy | 1436 ATCCAAATATGATGGGTTTTGAAACCCCTGTGGAACACAGAGGCTACTTCCACTCTTCC 1495 |
| Db | 330 ATCCAAATATGATGGGTTTTGAGCCCTGTGTAACAGAGATTACTTCCACCCACTTCC 389 |
| Qy | 1496 CTCGATATGCAAAATTAATTAAGGGAAGAAATGGTGAATCTTTTTCGAAGATTAATAG 1555 |
| Db | 390 CTCGCTATGCAAAATTAATTAAGGGAAGAAATGGTCAACTATTCTCAAGATTAATAG 449 |
| Qy | 1556 ATAGAAATAAAACCTCTGTGAGGTTGTGAATTTAAACAAATTAATGATCTGCTGGATT 1615 |
| Db | 450 ACAGAATAAAACCTGCTGGAAGTGTGAATTTACCAAACTTACATGATCTGCTGGATT 509 |
| Qy | 1616 TTTTCTGTGAATTTAGTGAACAGTCACATGTGTTTTCGAAGATCTCTGTTACAAACCA 1675 |
| Db | 510 TTTTCTGTGAATTTAGTGAACAGTCACCATGTGTTTTCGAAGATCTCTGTTACAAACCA 569 |
| Qy | 1676 CTTTCTCTGGTGGAATAACAAAAGGCTTTTGAATCTCATGCAAGACATGTTGAAAG 1735 |
| Db | 570 CTTTCTCTGGTGGAATAACAAAAGGCTTTTGAATCTCATGCAAGACATGTTGAAAG 629 |
| Qy | 1736 ATGCATCTGGTCTTTTGTGATGATCCCTCGAGTCTTTCCCGCAAGTGCTACCTATATA 1795 |
| Db | 630 ATGCTCTGGGTCCTTCGTGAG--TGCTCGGTGCTCTCCCGCAAGTGTGCTCATATA 687 |
| Qy | 1796 TAATCACCAGCTAAGGACTGTATGACTCCTTTGTTACTCACATGTGTTCGGCAATCTG 1855 |
| Db | 688 TAATCACCAGCTAAGGACTGTATGACTCCTTTGTTACTCACATGTGTTCGGCGGTTTG 747 |
| Qy | 1856 TAGTCTTATTCAGATCCATGACATACAGGCTCGACAGAGATTAAGCTTGTGTATAT 1915 |
| Db | 748 TAGTCTTATTCAGATCCATGAGCATTAACAGGCTCGGAGAGACAAAGCTTGTGTATAT 807 |
| Qy | 1916 TCTTGAGGAATTTGCCACCTTTCAGAGATGAGTTTATGACATTTTATTTTATAGGCAG 1975 |
| Db | 808 TCTTGAGGAGTTTGTCTACCTTTCAGATGAGT-----AACAAATTTTTCAGGCGAG 862 |
| Qy | 1976 AAGGTTGATCGAGCGCTTTCACACCATGCTGTTGAAACAGGAACCCCAAGCAACATTTG 2035 |
| Db | 863 AAGGTTGATCGAGCGCTTTCATCTATGTTGTTGAAACAGGAGCTTCAGAGACAACTCTG 922 |
| Qy | 2036 GCCTGTTTAGGTACCTGGGTCTTTTACCATAACCTTCGCATTTATGATACAGTACCTCTTA 2095 |

| | | | | | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|-----------------------------|--|--|
| Db | 923 | GCCTGCTTAGGAACCTGGGTTCTTTACCAATAACCTCGGATTAATGATCAGTATCTGCTC | 982 | | |
| Qy | 2096 | AGTGGCTTTGAATTGGAACTCTACAGTATGCACAGTACTATTACATATATGGT | 2150 | | |
| Db | 983 | AGTGGCTTTGAGCTGGAGCTGTACAGCATGCATGAGTACTACTACTACTCTGTT | 1037 | | |
| RESULT 7 | | | | | |
| EG547889 | | | | | |
| LOCUS | 602576187F1 NIH_MGC_77 | 768 bp | mRNA linear EST 04-APR-2001 | | |
| DEFINITION | mRNA sequence. | | | | |
| ACCESSION | EG547889 | | | | |
| VERSION | EG547889.1 | GI:13546554 | | | |
| KEYWORDS | EST. | | | | |
| SOURCE | Homo sapiens (human) | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | 1 (bases 1 to 768) | | | | |
| AUTHORS | NIH-MGC http://mgs.nci.nih.gov/. | | | | |
| TITLE | National Institutes of Health, Mammalian Gene Collection (MGC) | | | | |
| JOURNAL | Unpublished | | | | |
| COMMENT | Contact: Robert Strausberg, Ph.D. Email: cga@ra-remail.nih.gov Tissue Procurement: CLONTECH Laboratories, Inc. cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCM1543 row: n column: 21 High quality sequence stop: 738. | | | | |
| FEATURES | | | | | |
| source | 1..768 | | | | |
| | Location/Qualifiers | | | | |
| | /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:4704212" /lab_host="DH10B (T1 phage-resistant)" /clone_lib="NIH_MGC_77" | | | | |
| | /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site: 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcaattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGGGCGGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library." | | | | |
| BASE COUNT | 241 a | 152 c | 179 g 196 t | | |
| ORIGIN | | | | | |
| Query Match 21.1%; Score 653.6; DB 10; Length 768; Best Local Similarity 96.9%; Pred. No. 1.3e-107; Matches 698; Conservative 0; Mismatches 19; Indels 3; Gaps 3; | | | | | |
| Qy | 2149 | GTATCTCTCTGAATTCCTTTTACGCATGTTGATGTCAACATTTAGTGTGCGGATGGCTC | 2208 | | |
| Db | 1 | GTATCTCTCTGAATTCCTTTTACGCATGTTGATGTCAACATTTAGTGTGCGGATGGCTC | 60 | | |
| Qy | 2209 | TCAATGGCAGAGAAAGGATAATGGAAGCAGCAGAGGCGGTAGTAGTAAAAAAC | 2268 | | |
| Db | 61 | TCAATGGCAGAGAAAGGATAATGGAAGCAGCAGAGGCGGTAGTAGTAAAAAAC | 120 | | |
| Qy | 2269 | AAAGAAAAAAGAAAGTTTCCCATTTGAGCCGAGAGATCACATGAGCCAGCATATCA | 2328 | | |
| Db | 121 | AAAGAAAAAAGAAAGTTTCCCATTTGAGCCGAGAGATCACATGAGCCAGCATATCA | 180 | | |
| Qy | 2329 | GAACATGTGTCTGGAATGTTTAAACCATGGTAGCATTTGACATGACGACGCAAGTACG | 2388 | | |

Db 181 GAAATGTTGCTGGAATGTTTAAACCACTGATGACATTTGACATGGACGCAAGATACG 240
QY 2389 TAAACCGAAGTTTGAGCTGTATGATGAACAGTTTCGGTATGACACACAGTTTGTCTCCATT 2448
Db 241 TAAACCGAAGTTTGAGCTGTATGATGAACAGTTTCGGTATGACACACAGTTTGTCTCCATT 300
QY 2449 CAACAGTGTGATGACCCCGCCAGTGCACCTACTTACAGTTTCAAGGAAATGCTGACCT 2508
Db 301 CAACAGTGTGATGACCCCGCCAGTGCACCTACTTACAGTTTCAAGGAAATGCTGACCT 360
QY 2509 CAATAA-ATATAGCCCTCTCTCTAGTCTCTGAACTCTGAACTGATGTCGACCTAGTAAAGCACT 2567
Db 361 CAATAAATATAGCCCTCTCTCTAGTCTCTGAACTGATGTCGACCTAGTAAAGCACT 420
QY 2568 TTCAACAGCAGAAATATGATTTGAAATATTTCTTAACCCGACCATGAGGTTAATAGAA 2627
Db 421 TTCAACAGCAGAAATATGATTTGAAATATTTCTTAACCCGACCATGAGGTTAATAGAA 480
QY 2628 TTTTAAAGTTGCCAAAACCAACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAA 2687
Db 481 TTTTAAAGTTGCCAAGACCAACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAA 540
QY 2688 AGGAATCTAAAGTTCTCTCTGAAATTTGATTTCTGCTCATAAATATTTTCTGTTGTA 2747
Db 541 AGGAATCTAAAGTTCTCTCTGAAATTTGATTTCTGCTCATAAATATTTTCTGTTGTA 600
QY 2748 AACTTGTGTTGAGAGACTGGGGAGTGGCCAT-AAAGGGCAGAGTCTTCTTTGAGACC 2806
Db 601 AACTTGTGTTGAGAGACTGGGGAGTGGCCATCAACGGGAGAGTCTTCTTATCAGACC 660
QY 2807 CAACCTCTTAGAGGACATCACAGGCTCCACATCACGGGAAGTGAATGTTCTTTGG 2866
Db 661 CAATCTTAGAGGCAATTCACAGGCTCCA-ATCAGGGAAGTGAATGTTCTTGG 719

RESULT 8
LOCUS BQ965085
DEFINITION AGENCOURT_10052197 NIH_MGC_134 Mus musculus cdna clone
IMAGE:6509503 5', mRNA sequence.
ACCESSION BQ965085
VERSION BQ965085.1 GI:22380563
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-research.nih.gov
Tissue Procurement: Dr. David Rowe
cdna library Preparation: Invitrogen Corp
cdna library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL4076 row: o column: 08
High quality sequence stop: 632.
Location/Qualifiers
1. .957
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6509503"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_134"
/notes="Vector: pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2:

FEATURES
source

NotI; Cloned unidirectionally. Primer: Oligo dt. Average
insert size 1.7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH_MGC Library."
BASE COUNT 274 a 198 c 253 g 232 t
ORIGIN
Query Match 20.9%; Score 648; DB 13; Length 957;
Best Local Similarity 87.7%; Pred. No. 1.3e-106;
Matches 719; Conservative 0; Mismatches 100; Indels 1; Gaps 1;
QY 541 GGAGAGTAGCATAATGTTATGAAAGCTTCTGTAGATGATGACGATTCAGATGGGAG 600
Db 26 GCGCGGACGACATCATGGTTATGAAAGCTACTGTAGACGACGATGCTTCCGATGGGAG 85
QY 601 CTCAGTATGCGAGAAAATGAGAAAAGCAATACAAATCGGTGGACATTAACCAAGAT 660
Db 86 CTCGGGTCCCGGAAAATGAAAAGTAGCACAGCTGGTGGACATTAACCAAGAT 145
QY 661 TTTGAAGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCTATTGGT 720
Db 146 TTTGAAGATGCTTGTGAGAGCTGAAGTTGGGAGAACTGCTTCATGATAAGCTATTGGT 205
QY 721 CTTTGTGAAGCCTGCTGCTATTGAAATGATGATCCAGATGATGCTGCGATGAT 780
Db 206 CTTTGTGAAGCCTGCTGCTATTGAAATGATGATCCAGATGATGCTGCGATGAT 265
QY 781 GGAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAAGCTATCAAGGATGGCACTATT 840
Db 266 GGAACCAAGTTAATGAAAAGTTCTCAATTTTGAACAAAGCTATCAAGGATGGCACTATT 325
QY 841 AAAATTAAGATCTCACCTTGCCTGAACTGATAGGATTTATGGATACATGTTTGTCTGT 900
Db 326 AAAATTAAGAGCTCAGCTGCTGAACTGATAGGAATAATGGAACACCTGTTTCTGCTGT 385
QY 901 TTGATAAGCTGTTGAGAGGCCATTCAGTGGCAGACAGATTTTACGTCCTTTACATT 960
Db 386 TTGATCACATGCTCGAAGGCCATTCCTTGGCAGACAGATGTTTACGTCCTTTACATT 445
QY 961 CATATCTCAGACTTTTATAGAGATCCTGCTATGAAGGCTTTTGTCTCTGGAAATCTTGA 1020
Db 446 CATATCTCAGACTTTTATAGAGATCCTGCTATGAAGGCTTTTGTCTCTGGAAATCTTGA 505
QY 1021 ATCTGTGACATTCGAAGGGAAGAAAGTAAATAAAGCTGCTGTTTGTGAAGGGAAGATTTT 1080
Db 506 ATCTGTGACATTCGAAGGGAAGAAAGTAAATAAAGCTGCTGTTTGTGAAGGGAAGATTTT 565
QY 1081 CAGTCAATGACTTATGATTTTAAATGGCTAAAGTGTGACAGATCTTCAGTTACAGGC 1140
Db 566 CAGTCAATGACTTATGATTTTAAATGGCTAAAGTGTGACAGATCTTCAGTTACAGGC 625
QY 1141 ATGCTAAAAGATGTGGAGATGACATGCAAAAGAGATGAAAGAGTACTCGAAGTCGACAA 1200
Db 626 ATGCTAAAAGATGTGGAGATGATCTGCAAGGCGAGTAAAGTACTCGAAGTCGACAA 685
QY 1201 GGAGAAAGAGAGATCCGAAAGTTGAACTAGAACCAACCAATGTTTACAGTATTTCAGC 1260
Db 686 GGAGAAAGAGAGATCCGAGGTCGAACTAGAACCAACCAATGTTTACAGTATTTCAGC 745
QY 1261 AGAGTGAATTTTACTCTGTTGTTACTGACAGTCTTATAGCCCTTTACTAAGAAAGAGCC 1320
Db 746 AGAGTGAATTTTACTCTGTTGTTACTGACAGTCTTATAGCCCTTTACTAAGAAAGAGCC 804
QY 1321 AGTGTGTTGCAAGAGCTCAAAAATTTGATGTTTCAAGCAG 1360
Db 805 AGTGTGTTGCAAGAGCTCAAAAATTTGATGTTTCAAGCAG 844
RESULT 9
LOCUS A1790514/c
DEFINITION ul02b03 x1 Sugano mouse kidney m2ia Mus musculus cdna clone
IMAGE:2064845 3' similar to WP:23B12.4 CE14032 ;, mRNA sequence.
ACCESSION A1790514

| | |
|---------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| AI790514.1 | GI:5338230 |
| VERSION | |
| KEYWORDS | |
| SOURCE | Mus musculus (house mouse) |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. |
| REFERENCE | 1 (bases 1 to 828) |
| AUTHORS | Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson.R. |
| TITLE | The WashU-NCI Mouse EST Project 1999 |
| JOURNAL | Unpublished |
| COMMENT | Other_ESTRs: ul02b03.yl Contact: Marra M/WashU-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LInL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:994033 Possible reversed clone: similarity on wrong strand Seq primer: custom primer used High quality sequence stop: 509. Location/Qualifiers 1..828 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL" /db_xref="taxon:10090" /clone="IMAGE:2064845" /sex="female" /dev_stage="adult" /lab_host="DH10S" /clone_lib="Sugano mouse kidney mkia" /notes="Organ: Kidney; Vector: pME18S-FU3; Site_1: DraIII (CACTGTG); Site 2: DraII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCTACTGG], digested and cloned into distinct DraIII sites of the pm18S-FU3 vector (5' site CACTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCCTCTTAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGACA." BASE COUNT 208 a 197 c 177 g 240 t ORIGIN Query Match 20.7%; Score 641.4; DB 9; Length 828; Best Local Similarity 86.4%; Pred. No. 2e-105; Matches 716; Conservative 0; Mismatches 112; Indels 1; Gaps 1; QY 673 TGTCCGAGAATTAAAGTTGGAGAACACTCACTTCATGATAAGCTATTGTGCTCTTTTGAAGCC 732 DB 828 TGTCGAGAGCTGAAGTTGGAGACTGCTTCATTGATAAGCTGTTGTTCTTTTGAAGCC 769 QY 733 ATGTCTGCTATTGAATGATGCCAAGATCGATCGTGGCATGATTGGAACCAAGATT 792 DB 768 ATGTCTGCTATTGAATGATGATCTNTAAGATGATGATGCGGGTATGATCNGAACCAAGTG 709 QY 793 AATCGAAAAGTTCTCAATTTTGACACAGCTATCAAGATGGCACATTAATAATTAAGAT 852 DB 708 AATAGNAAAGTCTCAATTTTGANCAAGCTATCAAGGATGGCAC-ATTAAATTTAAGAC 650 QY 853 CTCACCTTGCTCAACTGATAGGGATTATGGATACATGTTTTTGCTGTTTGTATACGTGG 912 DB 649 CTGAGCTGCTGATGATGAGAAATTAAGACACCTGTTCTGCTGTTTGTATCATGG 590 |
| FEATURES | source |
| QY 913 TTAGAAGGCCATTCACTGGCACAGACAGTAGTATTTACGTGCTTTTACATTTTCAATAATCCAGAC | 972 |
| DB 589 CTGGAAGGCCATTCTTTGGCACAGACAGTGTCTTACGTGCTTTTACATTTTCAATAATCCGCAC | 530 |
| QY 973 TTTATAGAAGATCTCTGATGAAGGCTTTTGTCTCTGGATCTTTGAANAATCTCTGACATT | 103 |
| DB 529 TTTATAGAAGATCTCTGATGAAGGCTTTTGTCTCTGGATCTTTGAANAATCTCTGACATT | 470 |
| QY 1033 GCAGGAAAAAGTAATAAAGCTGCTGTTTGTGAAAGAGAAAGATTTTTCAGTCAATGACT | 109 |
| DB 469 GCAGGAAAAAGTAATAAAGCTGCTGTTTGTGAAAGAGAAAGATTTTTCAGTCAATGACT | 410 |
| QY 1093 TAGGATTTAAATGGCTAAACAGTGTGACAGATCTTCGAGTTTACAGCATGCTTAAAGAT | 115 |
| DB 409 TAGGATTTAAATGGCTAAACAGTGTGACAGATCTTCGAGTTTACAGCATGCTTAAAGAT | 350 |
| QY 1153 GTGAGATGATCATGCAAGAGTAAGAGTACTCGAGTCTGCAAGAGAGAGAGAGA | 121 |
| DB 349 GTGAGATGATCATGCAAGAGTAAGAGTACTCGAGTCTGCAAGAGAGAGAGAGA | 290 |
| QY 1213 GATCCAGAAGTTGAACCTAGAACACCACAAATGTTTAGCAGTATTTCAGCAGAGTAAATTT | 127 |
| DB 289 GATCCAGAAGTTGAACCTAGAACACCACAAATGTTTAGCAGTATTTCAGCAGAGTAAATTT | 230 |
| QY 1273 ACTGCTGTTTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAGAGACCAAGTGTGTTGCA | 133 |
| DB 229 ACCGAGTGTGCTTCAAGTGTCTTACTAAGAAAGAGACCAAGTGTGTTGCA | 170 |
| QY 1333 GAAGTCTAAAATTTGATGTTTCAAGCAGCAGATCTTCTTCTGCCATTTCATATTCATTG | 139 |
| DB 169 GAGCTCAGAAACTGATGTGTCCAGGGGGCAGACCTCTTCTTCTGCCATTTCACACTCATG | 110 |
| QY 1393 CATCATGGCATCCAGGCCAGAGATGATCTACAAAAGAGAGATCATCCAATTTATGATGGGT | 145 |
| DB 109 CACACGGCATCAAGGCTCCGAATGGCACTACCAAAGAGACCATCCAATTAATGGGT | 50 |
| QY 1453 TTTGAACCCCTTGTGAACACAGAGCTACTTTCACCTACCTTCCTCCGAT | 1501 |
| DB 49 TTTGAGCCCCCTTGTTAACAGAGATTACTTCCCCCGCTTCCCCTCGNT | 1 |
| RESULT 10 | |
| BM979511/c | |
| LOCUS | 616 bp mRNA linear EST 21-FEB-2000 |
| DEFINITION | UI-CF-DUI1-adr-h-18-0-UI.s1 UI-CF-DUI1 Homo sapiens cDNA clone |
| ACCESSION | BM979511 |
| VERSION | BM979511.1 GI:19600035 |
| KEYWORDS | EST. |
| SOURCE | Homo sapiens (human) |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| AUTHORS | 1 (bases 1 to 616) |
| TITLE | Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery |
| JOURNAL | Genome Res. 6 (9), 791-806 (1996) |
| MEDLINE | 97044477 |
| PUBMED | 8889548 |
| COMMENT | Contact: McCray, PB McCrays Lab University of Iowa 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA Tel: 319 356 4866 Fax: 319 356 7171 Email: paul-mccray@uiowa.edu Tissue Procurement: Dr. M. J. Welsh, University of Iowa cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research |

Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES

Location/Qualifiers

1..616
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-DUI-adr-h-18-0-UI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-DUI"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-DUI is a normalized cDNA library containing the
following tissue(s): Primary Lung Epithelial Cells The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dr primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GGCTGTAGGC.
TAG LIB=UI-CF-DUI
TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
TAG_SEQ=GGCTGTAGGC"

BASE COUNT 190 a 123 c 98 g 205 t

ORIGIN

Query Match 19.7%; Score 610; DB 12; Length 616;

Best Local Similarity 100.0%; Pred. No. 9.6e-100;

Matches 610; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2487 AGTTCAAGGAATGTCTGACCTCAATAATATAGCCCTCTCTCAGTCTCTGAACTGT 2546

Db 616 AGTTCAAGGAATGTCTGACCTCAATAATATAGCCCTCTCTCAGTCTCTGAACTGT 557

QY 2547 ATGTGGCAGCTAGTAAAGCACTTTCACAGGCAAAATGATATGGAAATATCTTAACC 2606

Db 556 ATGTGGCAGCTAGTAAAGCACTTTCACAGGCAAAATGATATGGAAATATCTTAACC 497

QY 2607 CGGACCATGAGGTTAATGAATTTTAAAGGTTGCCAAACCCCACTTTGGTGTATGAAGT 2666

Db 496 CGGACCATGAGGTTAATGAATTTTAAAGGTTGCCAAACCCCACTTTGGTGTATGAAGT 437

QY 2667 TATTGGCAGGAGGACACAAAAGGAATCTAAAGTTCCTCTCTGAATTTGATCTGTCTC 2726

Db 436 TATTGGCAGGAGGACACAAAAGGAATCTAAAGTTCCTCTCTGAATTTGATCTGTCTC 377

QY 2727 ATAAATATTTCTCTGTGTGAACCTTTGTTGAGAGACTGGGAGGTGGCCATAAAGG 2786

Db 376 ATAAATATTTCTCTGTGTGAACCTTTGTTGAGAGACTGGGAGGTGGCCATAAAGG 317

QY 2787 CGAGAGTCTCTTTTCAGACCCCACTCTTAGAGGGCAATCATCCAGGCTCCACATCACGG 2846

Db 316 CGAGAGTCTCTTTTCAGACCCCACTCTTAGAGGGCAATCATCCAGGCTCCACATCACGG 257

QY 2847 AAGTGAGATGGATTTCTTGGGTAAACAACCTATATAAGAAATCTTTAGTTTGACGCC 2906

Db 256 AAGTGAGATGGATTTCTTGGGTAAACAACCTATATAAGAAATCTTTAGTTTGACGCC 197

QY 2907 TTATATGACATGAATGAACAACTGCTGTTTAAAGTGGTTTATATGTTCCATGGAGAAA 2966

Db 196 TTATATGACATGAATGAACAACTGCTGTTTAAAGTGGTTTATATGTTCCATGGAGAAA 137

QY 2967 CTGGTCTTATGAATGCAATGATGAACGTTATATGTTTATATACAGATTTAATCACAAA 3026

Db 136 CTGGTCTTATGAATGCAATGATGAACGTTATATGTTTATATACAGATTTAATCACAAA 77

QY 3027 TCATTTTTTATGAATGATTGAGTAAAAATAGTGTTTATAAAGGTTAATAAAATTTCTTGAC 3086

Db 76 TCATTTTTTATGAATGATTGAGTAAAAATAGTGTTTATAAAGGTTAATAAAATTTCTTGAC 17

QY 3087 AAAAAA 3096

Db 16 AAAAAA 7

RESULT 11

CA977646 LOCUS

DEFINITION AGENCOURT 11295472 NIH MGC 164 Mus musculus cDNA clone

IMAGE:30146905 5', mRNA sequence.

ACCESSION CA977646

VERSION CA977646.1 GI:27510300

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 900)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: Dr. David Rowe and Dr. Mina

cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: NDA0063 row: f column: 02

High quality sequence stop: 655.

Location/Qualifiers

1..900

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="IMAGE:30146905"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC 164"

/note="Vector: pCMV-Sport6.1.cdb; Site 1: EcoRV; Site 2:

NotI; Non-normalized full-length enriched library from

pooled mouse embryonic limb, maxilla and mandible, day

10.5 and 11.5 (size selected for the 0.5-1 kb fragments)

Cloned directionally, priming method: Oligo-dr. cDNA

enrichment: >1k bp, Average insert size 1.8k bp. Priming

sequence: 5'GACTAGTTCTAGATCGAGCGGCCCTT 3'. Tissue

contributed by: David Rowe. Library constructed by ResGen,

Invitrogen Corp."

BASE COUNT 235 a 213 c 212 g 239 t 1 others

ORIGIN

Query Match 19.4%; Score 599.2; DB 14; Length 900;

Best Local Similarity 85.5%; Pred. No. 8.1e-98;

Matches 700; Conservative 0; Mismatches 94; Indels 25; Gaps 2;

QY 1535 ACTATTTGCAAGATTAAATAGATGAATAAAACHTGTCGTGAGGTTGTGAATTAACAA 1594

Db 3 ACTATTTCTCAAGATTAAATAGACAGATAAAACCTGCTGTGAAGTCGTGAACATTACCAA 62

QY 1595 ATTTACATTTGATCTCTCGATTTTTTCTGTGAATTTAGTGAACAGTCACCATGTGTTCTTT 1654

Db 63 ACTTACATCTGATCTCTCGATTTTTTCTGTGAATTTAGTGAAGTCACCATGTGTTCTTT 122

QY 1655 CAAGATCTCTGTTTACAAACCACTTTTCTCTGGTGGATTAACAAAAGGCTTTTGGAACTCATC 1714

Db 123 CAAGATCTCTGTTTACAAACCACTTTTCTCTGGTGGATTAACAAAAGGCTTTTGGCACTCATC 182

```
Qy 1715 TCATGCAAGCATGGTGAAGATGCACCTTCGGTCTTTTGTGATGATCTCTCCGAGTGCCTTC 1774
Db 183 TCATGCAAGCATGGTGAAGATGCCTTCGGTCTCTTCGTGAG--TCCTCCGTCCTC 240
Qy 1775 CCCAAGTGTACCTATATATATATATATATATATATATATATATATATATATATATATATAT 1834
Db 241 CCCAAGTGTGTCCCTATATATATATATATATATATATATATATATATATATATATATATATAT 300
Qy 1835 TCACCTGTCTCGGCATCTCTGAGTCTTATTTATGATCCATGATGATGATGATGATGATGATGATG 1894
Db 301 TCACCTGTCTCGGCATCTCTGAGTCTTATTTATGATCCATGATGATGATGATGATGATGATGATG 360
Qy 1895 GAGAGATAAGCTTGGTGCATATCTTTGAGGAATTTGCCACCTTGCAGGATGATGATGATGATGATG 1954
Db 361 GAGAGACAAGCTTGGTGCATATCTTTGAGGATTTGCTACCTTGAAGATG----- 410
Qy 1955 ATTTTATTTTATAGGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2014
Db 411 -----AGGCAGAGAAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 457
Qy 2015 GAACCCCAAGGCAACATTTGGCTCTTTAGGTACTCTGGTCTCTTTACCAATACCTTCGC 2074
Db 458 GAGCTCAGAGACAACATCTGGCTCTGTTAGAACCTGGTCTTTTACCAATACCTTCGC 517
Qy 2075 ATTATGATACAGTACCTTCTAAGTGGCTTTCAATTTGGAATCTTACAGTATGACGAGTAC 2134
Db 518 ATTATGATCAGTATCTCTCAGTGGCTTTGAGCTGGAGCTGTACAGCATGATGATGATGATGATG 577
Qy 2135 TATTACATATATTTGGTATCTCTGAAATCTTTACGATGATGATGATGATGATGATGATGATGATG 2194
Db 578 TACTACATCTACTGCTACCTCTCGAGTTCCTGTATGATGATGATGATGATGATGATGATGATGATG 637
Qy 2195 CGTCCGATGCTCTCAATCGCAGAGAAAGGATATGAGAGACAGACAGAAAGGCGGT 2254
Db 638 CGCCTCAGCGCTCTCAGATGCGAGAGAGAGATCATGGANGAACAGCAAGAGGCGCG 697
Qy 2255 AGTAGTAAAAAACAAGAAAAAAGAAAGATTCGCCCATTTGAGCCGAGAGATCACAAATG 2314
Db 698 AGCGGCCAAAAAACAAGAAAAAAGAAAGATTCGCCCATTTGAGCCGAGAGATCACAAATG 757
Qy 2315 AGCAAGCATATCAGAACATGCTGCTGGATGTTTAA 2353
Db 758 AGCAGGCTATCAGAACATGCTGCTGGGATGTTCAA 796
```

RESULT 12

```
BI602827
LOCUS 603247187F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5294804 5',
DEFINITION mRNA sequence.
ACCESSION BI602827
VERSION BI602827.1 GI:15495766
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLES NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11745 row: n column: 21
High quality sequence stop: 643.
```

FEATURES

```
source
1. .697
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5294804"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_96"
note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 208 a 116 c 185 g 188 t
ORIGIN
```

```
Query Match 19.3%; Score 598.8; DB 12; Length 697;
Best Local Similarity 99.5%; Pred. No. 9.9e-98;
Matches 611; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
```

```
Qy 459 GAGAGGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 518
Db 6 GAGAGGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 65
Qy 519 CGTGGCGGCGGGGTGACCGAGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 578
Db 66 CGTGGCGGCGGGGTGACCGAGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 125
Qy 579 TGATGACGATTCAGATGGGAGCTCAGTATGCCAGAAAAATGAGAAAAAGCAATACAAA 638
Db 126 TGATGACGATTCAGATGGGAGCTCAGTATGCCAGAAAAATGAGAAAAAGCAATACAAA 185
Qy 639 CTGGGTGACATACCAAGATTTTGAAGAGCTTGTGAGAAATTAAGTTGGGAGACT 698
Db 186 CTGGGTGACATACCAAGATTTTGAAGAGCTTGTGAGAAATTAAGTTGGGAGACT 245
Qy 699 ACTTCATGATAAGCTATTTGGTCTTTTGAAGCCATGCTGCTATTGAAATGATGGATCC 758
Db 246 ACTTCATGATAAGCTATTTGGTCTTTTGAAGCCATGCTGCTATTGAAATGATGGATCC 305
Qy 759 CAAGATGATGCTGGCATGATTGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACA 818
Db 306 CAAGATGATGCTGGCATGATTGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACA 365
Qy 819 AGCTATCAAGATGGCAGCTATTAATTAAGATCTCACCTTGCCTGACTGATAGGAT 878
Db 366 AGCTATCAAGATGGCAGCTATTAATTAAGATCTCACCTTGCCTGACTGATAGGAT 425
Qy 879 TATGATACATGTTTTTGTGTTGATAACCGTGTGTAGAGGCCATTCACCTGGCAGAC 938
Db 426 TATGATACATGTTTTTGTGTTGATAACCGTGTGTAGAGGCCATTCACCTGGCAGAC 485
Qy 939 AGTATTTACGTGCTTTACATTCATATATATATATATATATATATATATATATATATATATATAT 998
Db 486 AGTATTTACGTGCTTTACATTCATATATATATATATATATATATATATATATATATATATATAT 545
Qy 999 TTTTGTCTCTGGGATCTTGAATCTCTGATTCGAGGAGGAAAAAGTA-AATTAAGCTG 1057
Db 546 TTTTGTCTCTGGGATCTTGAATCTCTGATTCGAGGAGGAAAAAGTA-AATTAAGCTG 605
Qy 1058 CTGTCTCTGAAGAG 1071
Db 606 CTGTCTCTGAAGAG 619
```

RESULT 13

```
AW007583/c
LOCUS AW007583 598 bp mRNA linear EST 08-MAR-2000
```

```

DEFINITION wt02f11.x1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2506317 3',
            mRNA sequence.
ACCESSION  AW007583
VERSION     AW007583.1  GI:5856446
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 598)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,
            , Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arraying: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 715 Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 459.

FEATURES
            source
            1..598
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:2506317"
            /sex="pooled"
            /tissue_type="colon"
            /lab_host="DH10B"
            /clone_lib="NCI_CGAP Co3"
            /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
            polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
            was prepared from 12 pooled bulk tumor samples and primed
            with a Not I - oligo(dT) primer. Double-stranded cDNA was
            ligated to Eco RI adaptors (Pharmacia), digested with Not
            I and cloned into the Not I and Eco RI sites of the
            modified pT7T3 vector. Library went through one round of
            normalization."
BASE COUNT  189 a 122 c 98 g 188 t 1 others
ORIGIN
Query Match 19.3%; Score 597; DB 9; Length 598;
Best Local Similarity 99.8%; Pred. No. 2.1e-97;
Matches 597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2490 TCAAGGAATGTCGACCTCAATAATATAGCCCTCCTCAGTCTCTGAACTGTATG 2549
Db 598 TCAAGGAATGTCGACCTCAATAATATAGCCCTCCTCAGTCTCTGAACTGTATG 539
QY 2550 TGGGAGCTAGTAGCACTTCAACAGCGCAAAATGATATGGAAATATTCCTAACCCGG 2609
Db 538 TGGGAGCTAGTAGCACTTCAACAGCGCAAAATGATATGGAAATATTCCTAACCCGG 479
QY 2610 ACCATGAGTTATAGAAATTTAAAGGTTGCCAACCCCACTTTGGTTATGAAGTTAT 2669
Db 478 ACCATGAGTTATAGAAATTTAAAGGTTGCCAACCCCACTTTGGTTATGAAGTTAT 419
QY 2670 TGGCAGGAGACACAAAAAGGAATCTAAAGTTCTCTGAAATTTCTCTGCTCATAT 2729
Db 418 TGGCAGGAGACACAAAAAGGAATCTAAAGTTCTCTGAAATTTCTCTGCTCATAT 359
QY 2730 AATATTTTCTGTGTGAACCTGTTTGAAGAGACTGGGGAGTGGCCATAAAGGGCA 2789
Db 358 AATATTTTCTGTGTGAACCTGTTTGAAGAGACTGGGGAGTGGCCATAAAGGGCA 299
QY 2790 GAGTCTCTTTCAGACCCCACTCTTAGAGGGCACATCACCAGGCTCCACATCACGGGAAG 2849

```

```

Db 298 GAGTCTCTTTCAGACCCCACTCTTAGAGGGCACATCACCAGGCTCCACATCACGGGAAG 239
QY 2850 TGAGATGGAATTTCTGGTAACTCACTATTATAGGAATACTTTTAGTTTGACACCTTA 2909
Db 238 TGAGATGGAATTTCTGGTAACTCACTATTATAGGAATACTTTTAGTTTGACACCTTA 179
QY 2910 TATGACATGAATGAAACTGCTGTTTAAAGTGGTTTATTATGTTCCATGGAAGAACTG 2969
Db 178 TATGACATGAATGAAACTGCTGTTTAAAGTGGTTTATTATGTTCCATGGAAGAACTG 119
QY 2970 GTCTTATTGAATGCAATTCATGAACCTTATATGTTTATACAGATTTAATCACAATCA 3029
Db 118 GTCTTATTGAATGCAATTCATGAACCTTATATGTTTATACAGATTTAATCACAATCA 59
QY 3030 TTTTATTGAATGATGAGTGAATAGTGTATATATAGAGTTAATAAATTTCTTGACA 3087
Db 58 TTTTATTGAATGATGAGTGAATAGTGTATATATAGAGTTAATAAATTTCTTGACA 1

RESULT 14
LOCUS    BU058931
DEFINITION
ACCESSION BU058931
VERSION    BU058931
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            718 bp mRNA linear EST 26-AUG-2002
            UI-M-FR0-cam-e-22-0-UI-r1 NIH_BMAP_FR0 Mus musculus cDNA clone
            IMAGE:6414333 5', mRNA sequence.
            BU058931.1 GI:22499220
            EST.
            Mus musculus (house mouse)
            1 (bases 1 to 718)
            NIH-MGC http://mgi.nci.nih.gov/
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
Seq primer: pYX-5,
            Location/Qualifiers
            1..718
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6"
            /db_xref="taxon:10090"
            /clone="IMAGE:6414333"
            /tissue_type="whole brain"
            /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
            /lab_host="DH10B (T1 phage resistant)"
            /clone_lib="NIH_BMAP_FR0"
            /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
            Site 2: Not I; The library was constructed according
            Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
            1996. Denatured RNA was size fractionated on a 1% agarose
            gel. First strand cDNA synthesis was primed with oligo-dT
            primer containing a Not I site. Double strand cDNA was
            size selected according to mRNA size fraction, ligated
            with EcoR I adaptor, digested with NotI and then cloned
            directionally into pYX-Asc vector. The library tag
            sequence located between the Not I site and the polyA tail
            is ACCGAGAC. This library was created for the University
            Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the
            Developing Mouse Nervous System', supported by National
            Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,

```


Db 601 AGGAATCTAAAGTTCTCTGAA-TTGATTCTCTGCTCATA---ATATTCTGTGTGA 656
Qy 2748 AACTTGTTGAGAGAGACTGGGGAGG 2773
Db 657 AACTGTGAGAGAGACTGGGGAGTGG 682

Search completed: November 25, 2003, 02:50:11
Job time : 6201 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 24, 2003, 22:01:10 ; Search time 96.5 Seconds
(without alignments)
10184.818 Million cell updates/sec

Title: US-10-001-857-42

Perfect score: 5439

Sequence: 1 ttcttcacgaaactccagg.....attcttgacaaaaaaa 3096

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DBV=xlh
-Q=/cgn2.1/USPTO.spool/US10001857/runat_24112003_162854_9527/app_query.fasta.1.3271
-DB=A Geneseq 19Jun03 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US10001857 @CGN_1_1_0 @runat_24112003_162854_9527 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG-SUM=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseq 19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match % | Length | DB ID | Description |
|------------|--------|---------|--------|-----------|--------------------|
| 1 | 3796 | 69.8 | 733 | ABP52902 | Human lung speci |
| 2 | 3732.5 | 68.6 | 725 | AAOI19400 | Human molecule for |
| 3 | 1298.5 | 23.9 | 784 | ABG60398 | Drosophila melanog |
| 4 | 1237 | 22.7 | 239 | ABG69816 | Human secretory pr |
| 5 | 392 | 7.1 | 108 | AAU32414 | Novel human secret |
| 6 | 158.5 | 2.9 | 681 | ABB93650 | Herbicidally activ |
| 7 | 133.5 | 2.8 | 156 | ABG14149 | Novel human diagno |
| 8 | 146.5 | 2.7 | 120 | ABU91504 | Herbicidally activ |
| 9 | 146.5 | 2.7 | 120 | ABU52794 | Human transmembran |
| 10 | 146.5 | 2.7 | 574 | AAU52322 | WASP homolog prote |
| 11 | 146.5 | 2.7 | 574 | AAU52322 | Amino acid sequenc |
| 12 | 146 | 2.7 | 416 | ABG79529 | Human tumour suppr |
| 13 | 145.5 | 2.6 | 694 | ABG92421 | Herbicidally activ |
| 14 | 145 | 2.6 | 149 | AAU05346 | Human polypeptide |
| 15 | 145 | 2.6 | 559 | AAU52317 | Human Scari protei |
| 16 | 145 | 2.6 | 559 | AAU52317 | Amino acid sequenc |
| 17 | 145 | 2.6 | 559 | AAU52317 | Human pablo (pro-a |
| 18 | 145 | 2.6 | 559 | AAU52317 | Human WAVE1 (WAVE/ |
| 19 | 144 | 2.6 | 1134 | AAU24341 | Human lung-specifi |
| 20 | 143.5 | 2.6 | 1938 | ABP76682 | Streptomyces virid |
| 21 | 143 | 2.6 | 598 | ABG14000 | Novel human diagno |
| 22 | 143 | 2.6 | 780 | ABU00297 | Human novel polype |
| 23 | 143 | 2.6 | 1248 | AAU13464 | Human diaphanous p |
| 24 | 143 | 2.6 | 1938 | ABG98398 | Streptomyces virid |
| 25 | 141.5 | 2.6 | 551 | ABG93105 | S. cerevisiae BAX- |
| 26 | 141.5 | 2.6 | 560 | ABG21040 | Novel human diagno |
| 27 | 141.5 | 2.6 | 5877 | AAE34702 | Human mucin (MUC-1 |
| 28 | 141 | 2.6 | 191 | AAU29836 | Arabidopsis thalia |
| 29 | 141 | 2.6 | 199 | AAU29836 | Arabidopsis thalia |
| 30 | 140.5 | 2.6 | 1938 | ABP76680 | Streptomyces virid |
| 31 | 140 | 2.5 | 541 | AAU37118 | Mammalian Ena (Men |
| 32 | 140 | 2.5 | 783 | AAU37151 | Mouse neural Mena+ |
| 33 | 140 | 2.5 | 787 | AAU37152 | Mouse neural Mena+ |
| 34 | 140 | 2.5 | 802 | AAU37153 | Mouse neural Mena+ |
| 35 | 140 | 2.5 | 802 | AAU09139 | Mammalian enabled |
| 36 | 139.5 | 2.5 | 1243 | AAE34865 | Human kinases and |
| 37 | 139.5 | 2.5 | 1243 | ABP71620 | Human WNK4 protein |
| 38 | 139 | 2.5 | 214 | AAU86913 | Cotton fiber-speci |
| 39 | 138.5 | 2.5 | 250 | AAU67470 | Np70 protein carbo |
| 40 | 138.5 | 2.5 | 641 | AAU82327 | Human Npw388p1 tra |
| 41 | 138.5 | 2.5 | 641 | AAU67469 | Np70 protein seque |
| 42 | 138.5 | 2.5 | 641 | AAU72165 | Human RNA metaboli |
| 43 | 138.5 | 2.5 | 641 | AAU47514 | NpwBP. Homo sapie |
| 44 | 138.5 | 2.5 | 647 | AAU53462 | Human colon cancer |
| 45 | 138.5 | 2.5 | 903 | ABP65235 | Hypoxia-regulated |

ALIGNMENTS

RESULT 1

ABP52902

ID ABP52902 standard; Protein; 733 AA.

XX ABP52902;

XX 05-NOV-2002 (first entry)

DE Human lung specific protein sequence SEQ ID NO:145.

XX Human; lung; lung specific nucleic acid; LSNA; lung specific protein;
XX LSP; cytosolic; gene therapy; vaccine; metastasis; lung cancer;
XX squamous cell carcinoma.

XX Homo sapiens.

XX WO200264788-A2.

XX

PD 22-AUG-2002.
XX
XX
XX 20-NOV-2001; 2001WO-US45080.
XX
XX 20-NOV-2000; 2000US-252054P.
XX
XX (DIAD-) DIADEXUS INC.
XX
XX Macina RA, Recipon H, Chen S, Sun Y, Liu C;
XX WPI; 2002-657601/70.
DR
XX
XX New lung specific nucleic acid useful in gene therapy or as vaccines
XX for treating lung cancer (e.g. squamous cell carcinoma) or
XX non-cancerous lung diseases, as well as for diagnosing, monitoring or
XX staging these diseases
XX
XX Claim 11; Page 239-242; 282pp; English.
XX
XX The present invention describes an isolated lung specific nucleic acid
XX (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid
XX sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965;
XX (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp),
XX given in ABQ5262 to ABQ5376; (c) selectively hybridizes to (a) or (b);
XX or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific
XX protein (LSP) sequences have cytotatic activity and can be used in gene
XX therapy and vaccines. LSNA and LSPs are useful for diagnosing and
XX monitoring the presence and metastases of lung cancer in a patient. An
XX antibody that specifically binds to an LSP can be used for determining
XX the presence of an LSP in a sample, as well as for treating a patient
XX with lung cancer, particularly by inducing an immune response against
XX the lung cancer cell expressing the LSNA or LSPs. In particular, these
XX LSNA and LSPs are useful for identifying, diagnosing, monitoring,
XX staging, imaging and treating lung cancer (e.g. squamous cell carcinoma)
XX and non-cancerous disease states in lung.
XX
XX SQ Sequence 733 AA;

Alignment Scores:
Pred. No.: 0 Length: 733
Score: 3796.00 Matches: 732
Percent Similarity: 99.53% Conservative: 0
Best Local Similarity: 99.53% Mismatches: 1
Query Match: 69.79% Indels: 3
DB: 23 Gaps: 0

US-10-001-857-42 (1-3096) x ABP52902 (1-733)
QY 556 ATGGTTATGAAGCTTCTGTAGATGACGATTCAGGATGGAGCTGATGCGAGAA 615
Db 1 MetValMetLysAlaSerValAspAspAspSerGlyTrpGluLeuSerMetProGlu 20

QY 616 AAAATGGAGAAAAGCAATACAACTGGGGGAGCATTTACCAAGATTTTGAAGAAGCTGT 675
Db 21 LysMetGluLysSerAsnThrAsnTrpValAspIleThrGlnAspPheGluGluAlaCys 40

QY 676 CGAGATTAAGTTGGGAACTACTTCATGATAAGCTATTTGGTCTTTTGAAGCCATG 735
Db 41 ArgGluLeuLysLeuGlyGluLeuLeuHisAspLysLeuPheGlyLeuPheGluAlaMet 60

QY 736 TCTGCTATTGAATGATGATCCAGATGGATGCTGGCATTTGGAACCAAGTTAAT 795
Db 61 SerAlaIleGluMetMetAspProLysMetAspAlaGlyMetIleGlyAsnGlnValAsn 80

QY 796 CGAAAAGTTCTCAATTTTCAACAAGCTATCAAGGATGGCACTATTAAATTAAGATCTC 855
Db 81 ArgLysValLeuAsnPheGluGlnAlaIleLysAspGlyThrIleLysIleLysAspLeu 100

QY 856 ACCTTGCTGAAGTGAAGGATTTATGATACATGTTTGTGTTTGTATTAACGGTGA 915
Db 101 ThrLeuProGluLeuIleGlyIleMetAspThrCysPheCysCysLeuIleThrTrpLeu 120

QY 916 GAAGCCATTCAGTGGCAGACAGATTTTACGTGCTTTATCATTAATCCAGACTTT 975

Db 121 GluGlyHisSerLeuAlaGlnThrValPheThrCysLeuTyrIleHisAsnProAspPhe 140
QY 976 ATAGAAGATCCTGCTATGAAGCTTTTCTCTGGGATCTTGAATACTGTGACATTCGA 1035
Db 141 IleGluAspProAlaMetLysAlaPheAlaLeuGlyIleLeuLysIleCysAspIleAla 160
QY 1036 AGGAAAAAGTAAATAAAGCTGCTGTTTTCAGAGAGGAAGATTTTCAGTCAATGACTTAT 1095
Db 161 ArgGluLysValAsnLysAlaValPheGluGluAspPheGlnSerMetThrThr 180
QY 1096 GGATTTAAATGGCTAAAGTGTGACAGATCTTCAGATTACAGGATCTCTAAAAGATGTG 1155
Db 181 GlyPheLysMetAlaAsnSerValThrAspLeuArgValThrGlyMetLeuLysAspVal 200
QY 1156 GAGATGACATGCAAGAGAGAGTAAAGAGTACTCGAAGTCCGACAGGAGAGAGAGAGAT 1215
Db 201 GluAspAspMetGlnArgValLysSerThrArgSerArgGlnGlyGluArgAsp 220
QY 1216 CCAGAAGTTGAACAGACACCAATGTTTAGCAGTATTTCAGCAGAGTGAATTTACT 1275
Db 221 ProGluValGluLeuGluHisGlnCysLeuAlaValPheSerArgValLysPheThr 240
QY 1276 CGTGTGTTACTGACAGTCTTATAGCCTTTACTAAGAAAGAGACACAGTCTGTTCAGAA 1335
Db 241 ArgValLeuLeuThrValLeuIleAlaPheThrLysLysGluThrSerAlaValAlaGlu 260
QY 1336 GCTCAAAATGATGGTTCAAGCAGCAGATCTTCTTCTGCATTCATTAATTCATTGAT 1395
Db 261 AlaGlnLysLeuMetValGlnAlaAlaAspLeuLeuSerAlaIleHisAsnSerLeuHis 280
QY 1396 CATGCATCCAGGCCAGATGATACATAAAAAGGAGATCATCCAAATATGATGGGTTTT 1455
Db 281 HisGlyIleGlnAlaGlnAsnAspThrThrLysGlyAspHisProIleMetMetGlyPhe 300
QY 1456 GAAACCCCTTGTGAACAGAGGCTACTTCCACTTCCCTCCGATATGCAAAATAATTT 1515
Db 301 GluProLeuValAsnGlnArgLeuLeuProThrPheProArgTrpAlaLysIleIle 320
QY 1516 AAAAGGGAAGAAATGGTGAATCTTTTTCAGAGATTAATAGATAGATAAAATCTCTGT 1575
Db 321 LysArgGluGluMetValAsnTrpPheAlaArgLeuIleAspArgIleLysThrValCys 340
QY 1576 GAGGTGTGAATTTAAACAAATTTACATTTCTCTCTGGATTTTCTGTGAATTTAGTGAA 1635
Db 341 GluValValAsnLeuThrAsnLeuHisCysIleLeuAspPhePheCysGluPheSerGlu 360
QY 1636 CAGTCACCATGTGTTCTTTCAAGATCTCTGTACAAACCACTTTCTCTGTGTGATACAAA 1695
Db 361 GlnSerProCysValLeuSerArgSerLeuLeuGlnThrThrPheLeuValAspAsnLys 380
QY 1696 AAGGTCTTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACTTCGGTCTTTTGT 1755
Db 381 LysValPheGlyThrHisLeuMetGlnAspMetValLysAspAlaLeuArgSerPheVal 400
QY 1756 AGATCCTCCGATGCTTTTCCCAAGTCTACCTATAATAATACACAGGCTAAGGACT 1815
Db 401 -AspProPro-ValLeuSerProLysCysTyrLeuTyrAsnAsnHisGlnAlaLysAspC 420
QY 1816 GTATCGACTCTTGTGTTTACTCTGCTGTGTCGCAATCTGTAGTCTTATTGATCCATG 1875
Db 420 ysIleAspSerPheValThrHisCysValArgProPheCysSerLeuIleGlnIleHisG 440
QY 1876 GACATAACAGGGCTCGACAGAGATAAGCTTGGTCATATCTTCAGGAATTTGGCACCT 1935
Db 440 lyHisAsnArgAlaArgGlnArgAspLysLeuGlyHisIleLeuGluGluPheAlaThrL 460
QY 1936 TGCAGATGAGTTTATGACATTTTATTATTAATAGCA-GAGAAGGTTTATGACAGCGCTTC 1994
Db 460 euGlnAspGluPheMetThrPheTyrPheAsnArgAlaGluLysValAspAlaLeuLeuH 480
QY 1995 ACACCATCTGTTGAAACAGGAAACCCCAAGCAACATTTGGCCTGTTTAGTACTGGG 2054

```

Db      480  isThrMetLeuLeuLysGlnGluProGlnArgGlnHisLeuAlaCysLeuGlyThrTrpV 500
QY      2055  TCCTTTACCAATAACCTTCGCATTATGATACAGTACCTTCTAAGTCGGCTTGAATTGGAC 2114
Db      500  alLeuTyrHisAsnLeuArgileMetileGlnTyrLeuLeuSerGlyPheGluLeuGluL 520
QY      2115  TCTACAGTATGCACGAGTACTATTACATATATTCGTATCTCTCTCAATTCCTTTACGCAT 2174
Db      520  euTyrSerMetHisGluTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 540
QY      2175  GGTGATGTCAACATTTAGTTCGTCGCGATGCTCTCAATGCGCAGAGAAAGGATAATGG 2234
Db      540  rpLeuMetSerThrLeuSerArgAlaAspGlySerGlnMetAlaGluGluArgileMetG 560
QY      2235  AAGAGCAGCAAAAGCGGTAGTAAATAAACAAGAAAAAAGAAAAAGAACTTCGCCCAT 2294
Db      560  luGluGlnGlnLysGlyArgSerSerLysLysThrLysLysLysLysValArgProL 580
QY      2295  TGAGCCGAGAGATCACATATGACCAAGCATATCATGACATGCTGCTCGAATGTTTAAA 2354
Db      580  euSerArgGluileThrMetSerGlnAlaTyrGlnAsnMetCysAlaGlyMetPheLysT 600
QY      2355  CCATGGTAGCATTTGACATGACGCGCAAGTACGTAAACCGAAGTTTGAGCTTGATAGTG 2414
Db      600  hrMetValAlaPheAspMetAspGlyLysValArgLysProLysPheGluLeuAspSerG 620
QY      2415  AACAAAGTTCGGTATGAACACAGGTTTGCTCCATTCAACAGTGTGATGACCCCGCCAG 2474
Db      620  luGlnValArgTyrGluHisArgPheAlaProPheAsnSerValMetThrProProPro 640
QY      2475  TGCCTACTTACAGTCAAGGAATGTCTGACCTCAATAATATAGCCCTCTCTCTCAGT 2534
Db      640  alHisTyrLeuGlnPheLysGluMetSerAspLeuAsnLysTyrSerProProProGlnS 660
QY      2535  CTCCTGAACTGTATGTGCACCTAGTAAAGCACTTCAACAGCGAAAAATGATATGGAAA 2594
Db      660  erProGluLeuTyrValAlaAlaSerLysHisPheGlnGlnAlaLysMetileLeuGluA 680
QY      2595  ATATTCTCTAACCCGACCATGAGTTAATAGAAATTTTAAAGTTGCCAAACCCCACTTTG 2654
Db      680  snlleProAsnProAspHisGluValAlaAsnAlaGlleLeuLysValAlaLysProAsnPheV 700
QY      2655  TGGTTATGAAGTTATTGGCAGGAGACACAAAAAGGAATCTAAAGTTCTCTCTCTGAATTG 2714
Db      700  alValMetLysLeuLeuAlaGlyGlyHisLysLysGluSerLysValProProGluPheA 720
QY      2715  ATTTCTCTGCTCATAAATATTTCTCTGTTGTGAACCTTGT 2755
Db      720  spPheSerAlaHisLysTyrPheProValValLysLeuVal 733

RESULT 2
AAO19400
ID  AAO19400 standard; Protein; 725 AA.
XX  AAO19400;
XX  AAO19400;
XX  10-DEC-2002 (first entry)
XX  Human molecule for disease detection and treatment protein #3.
XX  Human; molecule for disease detection and treatment; MDDT; gene therapy;
KW  cytosatic; antiarteriosclerotic; hepatotropic; anti-HIV; anti-allergic;
KW  anti-inflammatory; antiasthmatic; cerebroprotective; nootropic;
KW  neuroprotective; antiparkinsonian; cardiant; antianginal.
XX  OS  Homo sapiens.
XX  PN  WO200270709-A2.
XX  PD  12-SEP-2002.
XX  PF  08-FEB-2002; 2002WO-US03709.
XX

```

```

PR      09-FEB-2001; 2001US-268117P.
PR      15-FEB-2001; 2001US-269618P.
PR      23-FEB-2001; 2001US-271118P.
PR      07-MAR-2001; 2001US-274486P.
PR      09-MAR-2001; 2001US-274436P.
PR      28-NOV-2001; 2001US-334229P.
PR      01-FEB-2002; 2002US-353284P.
XX  (INCY-) INCYTE GENOMICS INC.
XX
Lal PG, Baughn MR, Yao MG, Walia NK, Elliot VS, Xu Y;
PI Honchell CD, Yue H, Ding L, Gietzen KJ, Ison CH, Lu DAM;
PI Hafalia AUA, Ghandi AR, Thangavelu K, Sanjanwala MM, Tang YN;
PI Rankumar J, Griffin JA, Swarnaker A, Azimzai Y, Sapperstein SK;
PI Buford N, Lee EA, Lu Y, Iran UK, Marquis JP;
XX
WPI; 2002-713453/77.
DR N-PSDB; AAL49929.
XX
New human molecules for disease detection and treatment (MDDT), useful
PT for diagnosing, treating and preventing diseases or conditions
PT associated with the aberrant MDDT expression, e.g. cancer, AIDS,
PT asthma, diabetes, hepatitis -
XX
PS Claim 1; Page 128-130; 177pp; English.
XX
The present invention relates to human proteins and coding sequences of
CC molecules for disease detection and treatment MDDT. The sequences can be
CC used in the treatment of diseases associated with the decreased
CC expression or overexpression of MDDT, such as cell proliferative (cancer,
CC atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS,
CC allergies, Addison's disease, asthma), developmental (dwarfism, renal
CC tubular acidosis), neurological (e.g. stroke, Parkinson's disease,
CC epilepsy) and cardiovascular (congestive heart failure, myocardial
CC infarction, angina pectoris) disorders. The present sequence is a protein
CC of the invention.
XX
SQ Sequence 725 AA;
XX
Alignment Scores:
Pred. No.: 0 Length: 725
Score: 3732.50 Matches: 723
Percent Similarity: 98.23% Conservative: 0
Best Local Similarity: 98.23% Mismatches: 2
Query Match: 68.62% Indels: 11
DB: 23 Gaps: 1
US-10-001-857-42 (1-3096) x AAO19400 (1-725)
QY 556 ATGGTTATGAAGCTTCTGTAGATGATGACGATTTCAGATGGAGCTCAGTATGCCAGAA 615
Db 1 MetValMetLysAlaSerValAspAspAspSerGlyTyrGluLeuSerMetProGlu 20
QY 616 AAAATGGAGAAAGCAATACAAACTGGGTGCACATTACCCCAAGATTTTGAAGAAGCTTGT 675
Db 21 LysMetGluLysSerAsnThrAsnTyrValAspIleThrGlnAspPheGluGluAlaCys 40
QY 676 CGAAGATTAAGTTGGGAGAACTACTTCATGATTAAGCTATTGTGCTCTTTTGAAGCCCATG 735
Db 41 ArgGluLeuLysLeuGlyGluLeuLeuHisAspLysLeuPheGlyLeuPheGluAlaMet 60
QY 736 TCTGCTATTGAATGATGGATCCCAAGATGATGCTGGCATGATGGAAACCAAGTTAAT 795
Db 61 SerAlaIleGluMetMetAspProLysMetAspAlaGlyMetIleGlyAsnGlnValAsn 80
QY 796 CGAAAAGTCTCAATTTTGAACAAAGCTATCAAGATGGCACTATTAAATTAAGATCTC 855
Db 81 ArgLysValLeuAsnPheGluGlnAlaIleLysAspGlyThrIleLysIleLysAspLeu 100
QY 856 ACCTTGCCCTGAATGATAGGATATGATACATGTTTTCGTCTTTTGATAACGTGGTTA 915
Db 101 ThrLeuProGluLeuIleGlyIleMetAspThrCysPheCysLeuIleThrTripleu 120

```

| | | | |
|----|------|----------------------------------------------------------------|------|
| QY | 916 | GAAGGCCATTCACTGGCACAGACAGTATTTACCTGCTTTACATTCATTAATCCAGACTTTT | 975 |
| Db | 121 | GluGlyHisSerLeuAlaGlnThrValPheThrCysLeuTyzirileHisAsnProAspPhe | 140 |
| QY | 976 | ATAGAGATCCTGCTATGAAGGCTTTTGCCTCGGAATCTTGAAATCTGTGCACATTGCA | 1035 |
| Db | 141 | IleGluAspProAlaMetLysAlaPheAlaLeuGlyIleLeuLysIleCysAspIleAla | 160 |
| QY | 1036 | AGGGAATAAATAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTCAATGACTTAT | 1095 |
| Db | 161 | ArgGluLysValAsnLysAlaAlaValPheGluGluAspPheGlnSerMetThrTyr | 180 |
| QY | 1096 | GGATTTAAATGCTTAACGTGTGCACAGATCTTCGAGTTACAGCATGCTTAAAGATGCG | 1155 |
| Db | 181 | GlyPheLysMetAlaAsnSerValThrAspLeuArgValThrGlyMetLeuLysAspVal | 200 |
| QY | 1156 | GAGGATCACATGCAAGAGAGAGTAAAGAGTACTCGAAGTCACAAAGAGAGAAAGAGAT | 1215 |
| Db | 201 | GluAspAspMetGlnArgArgValLysSerThrArgSerArgGlnGlyGluArgAsp | 220 |
| QY | 1216 | CCAGAAGTTGAACTAGAACACCAACATGTTTAGCAGTATTCAGCAGAGTGAATTTACT | 1275 |
| Db | 221 | ProGluValGluLeuGluHisGlnGlnCysLeuAlaValPheSerArgValLysPheThr | 240 |
| QY | 1276 | CGTGTGTTTACTGACAGTGCCTTATAGCCCTTTACTAAGAAAGAGACCACTGCTGTGCAGAA | 1335 |
| Db | 241 | ArgValLeuLeuThrValLeuIleAlaPheThrLysLysGluThrSerAlaValAlaGlu | 260 |
| QY | 1336 | GCTCAAAAATTGATGGTTCAAGACGACAGATCTCTTTCTGCCATTATTAATCATTTGCAT | 1395 |
| Db | 261 | AlaGlnLysLeuMetValGlnAlaAlaAspLeuLeuSerAlaIleHisAsnSerLeuHis | 280 |
| QY | 1396 | CATGCGATCCAGGCCCAAGATGATACTACAAAAGAGATCATCCAATATATGATGGGTTTT | 1455 |
| Db | 281 | HisGlyIleGlnAlaGlnAsnAspThrThrLysGlyAspHisProIleMetMetGlyPhe | 300 |
| QY | 1456 | GAACCCCTTGTGAACACGAGAGCTACTTCCACCTACCTTCCCTCGATATGCAAAAATAATT | 1515 |
| Db | 301 | GluProLeuValAsnGlnArgLeuLeuProProThrPheProArgTyrAlaLysIleIle | 320 |
| QY | 1516 | AAAGGGAGAAATGGTGAACATAATTTGCAAGATTAATAGATAGAAATAAAATGCTGTG | 1575 |
| Db | 321 | LysArgGluGluMetValAsnTyrPheAlaArgLeuIleAspArgIleLysThrValCys | 340 |
| QY | 1576 | GAGGTGTGAATTTAAACAAATTTACATGTATCTCGGATTTTTTCTGTGAATTTAGTGAA | 1635 |
| Db | 341 | GluValValAsnLeuThrAsnLeuHisCysIleLeuAspPhePheCysGluPheSerGlu | 360 |
| QY | 1636 | CAGTCACCATGTGTTCTTTCAGATCTCTGTATCAAAACCATTTCCTGGTGGATACAAA | 1695 |
| Db | 361 | GlnSerProCysValLeuSerArgSerLeuLeuGlnThrThrPheLeuValAspAsnLys | 380 |
| QY | 1696 | AAGGTCTTTGAAATTCATCTCATGCAAGACATGTGTGAAGATGCATTCGGTCTTTTGTG | 1755 |
| Db | 381 | LysValPheGlyThrHisLeuMetGlnAspMetValLysAspAlaLeuArgSerPheVal | 400 |
| QY | 1756 | AGATCCTCCGAGTGTCTTCCCCCAAGTGCTACCTATATATATATATATATATATATAT | 1815 |
| Db | 401 | -SerProPro-ValLeuSerProLysCysTyrLeuTyrAsnAsnHisGlnAlaLysAspC | 420 |
| QY | 1816 | GTATCGACTCTTGTGTACTCACTGTGTTCGGCCATTCTGTGAGTCTTATTCAGATCCATG | 1875 |
| Db | 420 | ysIleAspSerPheValThrHisCysValAlaGProPheCysSerLeuIleGlnIleHisG | 440 |
| QY | 1876 | GACATAACAGGGCTCGACAGAGAGATAGCTTGGTTCATATTCTTTGAGGAATTTGCCACCT | 1935 |
| Db | 440 | lyHisAsnArgAlaArgGlnArgAspLysLeuGlyHisIleLeuLeuGluPheAlaThrL | 460 |
| QY | 1936 | TGCAGGATGAGTTTATGACATTTTATTTTAAATAGCCAGAGAGAGTTCATCCAGCGCTCA | 1995 |
| Db | 460 | eugInAsp-----Glu--AlaGluLysValAspAlaAlaLeuHi | 472 |
| QY | 1996 | CACCATGCTGTTGAAACAGGAACCCCCAAAGGCAACATTTGCCCTGTTTAGTACTCTGGT | 2055 |

| | | | |
|----------|------------------------------------------------------------------|---------------------------------------------------------------|------|
| Db | 472 | sThrMetLeuLeuLysGlnGluProGlnArgGlnHisLeuAlaCysLeuGlyThrIrpVa | 492 |
| Qy | 2056 | CCTTTACATAACCTTCGCATTTATGATACAGTACCTCTTAAGTGGCTTTCAATTTGGAAC | 2115 |
| Db | 492 | lLeuTyRHisLsnLeuArgileMetileGlnTyRLeuLeuSerGlyPheGluLeuGluLe | 512 |
| Qy | 2116 | CTACAGTATGCACGAGTACTATTACATATATTTGGTATCTCTCTGAATTCCTTTACGCATG | 2175 |
| Db | 512 | uTyRSerMetHisGlnuTyRtyrIleTyRTrpTyRLeuSerGluPheLeuTyRAlaIfr | 532 |
| Qy | 2176 | GTTCATCTCAACATTGAGTCGTGCCGATGCTCTCAATCGCAGAGAAAGGATAATGGGA | 2235 |
| Db | 532 | pLeuMetSerThrLeuSerArgAlaaspGlySerGlnMetAlaGluGluArgileMetGl | 552 |
| Qy | 2236 | AGAGCAGCAGAAAGCGGTAGTAGTAAAAAACAAGAAAAAAGAAAGTTTCGCCCAT | 2295 |
| Db | 552 | uGluGlnGlnLysGlyArgSerLysLysThrLysLysLysLysValArgProLe | 572 |
| Qy | 2296 | GAGCCGAGAGATCACATGAGCCACGACATATCAGAATCATGTGCTCGAATGTTTAAAC | 2355 |
| Db | 572 | uSerArgGluIleThrMetSerGlnAlaIfrGlnAsnMetCysAlaGlyMetPheLysTh | 592 |
| Qy | 2356 | CATGGTAGCATTTGCATGAGCGCAAGTACGTAAACCGAAGTTTCAGCTTGATAGTGA | 2415 |
| Db | 592 | xMetValAlaPheaspMetAspGlyLysValArgLysProLysPheGluLeuaspSerGl | 612 |
| Qy | 2416 | ACAAGTTCGGTATGAACACAGGTTTGCTCCATTCAACAGTGTGATGACCCGCGCCAGT | 2475 |
| Db | 612 | uGlnValArgTyRGlulHisArgPheAlaProPheAsnSerValMetThrProProProVa | 632 |
| Qy | 2476 | GCACCTACTACAGTTCAAGGAAATGTCGTACCTCAATAAATATAGCCCTCTCTCTCAGTC | 2535 |
| Db | 632 | lHisTyRLeuGlnPheLysGluMetSerAspLeuAsnLysTyRSerProProGlnSe | 652 |
| Qy | 2536 | TCCTGAACCTGTATGTCGAGCTAGTAGCACTTTTCAACAGCGCAAAAATGATATTGGAAA | 2595 |
| Db | 652 | xProGluLeuTyRValAlaAlaSerLysHisPheGlnGlnAlalysMetileLeuGluAs | 672 |
| Qy | 2596 | TATTCTTAACCCGACCATGAGGTTAATAGAAATTTTAAAGGTTGCCAAACCCCACTTTGT | 2655 |
| Db | 672 | nileProAsnProAspHisGluValAsnArgileLeuLysValAlalysProAsnPheVa | 692 |
| Qy | 2656 | GGTTATGAAGTATTGCGAGGAGCACACAAAAAGGAATCTAAAGTTCTCTCGAATTGA | 2715 |
| Db | 692 | lValMetLysLeuLeuAlaGlyGlyHisLysLysGluSerLysValProProGluPheAs | 712 |
| Qy | 2716 | TTTTCTCGCTCATAAATATTTTCTGTGTGGAACTTGTT | 2755 |
| Db | 712 | pPheSerAlaHisLysTyRPhProValValLysLeuVal | 725 |
| RESULT 3 | | | |
| ABB60388 | | | |
| ID | ABB60388 standard; Protein; 784 AA. | | |
| XX | ABB60388; | | |
| XX | | | |
| XX | | | |
| DT | 26-MAR-2002 (first entry) | | |
| XX | Drosophila melanogaster polypeptide SEQ ID NO 7956. | | |
| DE | | | |
| XX | Drosophila; developmental biology; cell signalling; insecticide; | | |
| KW | pharmaceutical. | | |
| XX | | | |
| OS | Drosophila melanogaster. | | |
| XX | | | |
| PN | WO200171042-A2. | | |
| XX | | | |
| PD | 27-SEP-2001. | | |
| XX | | | |
| PF | 23-MAR-2001; 2001WO-US09231. | | |
| XX | | | |
| PR | 23-MAR-2000; 2000US-191637P. | | |

```

PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL04491.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT Genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 7956; Zipp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLL16176-ABLL30511), expressed DNA
CC sequences (ABLO1840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 784 AA;
SQ

```

QY 2008 GAACAGGAACCCAAAGCAACATTGGCCCTGTTTAGTACCTGGTCTTTACCAATAA 2067
Db : : : : :
534 -----LysHisSerThrHisPheSerThrTrpValLeuTyrAsnCy 547
QY 2068 CTTGCGATATGATACAGTACTCTTAAGTGGCTTTGAATGGAACCTACAGTATGCA 2127
Db : : : : :
547 sPheArgAlaMetLeuIlePheLeuMetSerGlyPheGluLeuGluLeuTyrAlaValHi 567
QY 2128 CGAGTACTATTACATATATTGTTATCTCTCTGAATTCCTTTACGCCATGGTTGATGCAAC 2187
Db : : : : :
567 sGluPheLeuTyrIleTyrTrpPyrProTyrGluPheLeuIleGlyPheLeuValSerAl 587
QY 2188 ATTGAGTCGTGGCGTCTCAAAATGGCAGAGGAAGATAATGGAAGAGCAGCAGAA 2247
Db : : : : :
587 aLeuThrArgThrGluAsnIleLeuLeuAlaGlnGluTyrAlaGluHisGlnSerIy 607
QY 2248 AGCCCGTAGT-----AGTAAAAACAAGAA 2274
Db : : : : :
607 sThrGlnSerGlyGlySerGlyAlaAlaLysAsnArgLysAlaAlaLysProLysLysAs 627
QY 2275 AAAAAAGAAAGTTGCGCCCATTTGAGCGAGAGATCAACAATGAGCCAGCATATCAGAACAT 2334
Db : : : : :
627 nLysLysThrGlnArgProTyrArgAlaGluIleValPheTyrHisAlaLeuLeuSerIe 647
QY 2335 GTGTGCTGGAATGTTTAAACCATGTGATGACATGGACGGCAAGTACGTAAAC 2394
Db : : : : :
647 uCysGlyGlyMetTyrLysAlaMetGlyAlaLeuThrLysAspGlyArgValArgLeuPr 667
QY 2395 GAAGTTTGAGTTGATAGTACACAGTTGCGTATGACACAGTTTCTCCATTCACAG 2454
Db : : : : :
667 oLeuSerLysPheAspAsnGluGluIleArgTyrAsnArgPheLeuProPheAlaIath 687
QY 2455 TGTGATGACCCGCCCGCCAGTGCACACTTACAGTTCAAGAAATGCTGAC---CTCAA 2511
Db : : : : :
687 rLeuThrSerProProValSerTyrAlaGluPheLysAsnValArgGluHisMetMe 707
QY 2512 TAAATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGGCAGCTAGTAGACATTTCA 2571
Db : : : : :
707 tArgHisSer-----ValGluAspLeuTyrThrTyrAlaAlaLysHisPheAs 723
QY 2572 ACAGGCAAAATGATATTGAAATATCTCTACCCGACCATAGGTTTAATAGATT 2631
Db : : : : :
723 pGlnAlaArgAsnValLeuGluSerIleGlnAsnProAspGlnGluMetLeuAspLeuLe 743
QY 2632 AAAGTTGCAACCAACCTTTCTGGTATGAAAGTTATTGGCAGGAGGACACAAAAGCA 2691
Db : : : : :
743 uGlnIleAlaArgThrAsnPheValValMetAsnValLeuAlaArgGlyHisGlnLysGI 763
QY 2692 ATCTAAAGTTCCTCGAATTGATTCTCTGCTCTAATAATTTCTGTTGTGAACT 2751
Db : : : : :
763 uValLysArgGlnProGluPheAspPheSerLysHisSerTyrPheProIleIleLysLe 783
QY 2752 T 2752
Db 783 u 783
RESULT 4
ABG69816
ID ABG69816 standard; Protein; 239 AA.
XX AC ABG69816;
XX AC ABG69816;
XX 21-OCT-2002 (first entry)
XX Human secretory protein #7.
XX Human; secretory polypeptide; SPTM; actinic keratosis; arteriosclerosis;
KW bursitis; cirrhosis; hepatitis; polycythaemia vera; anaemia; psoriasis;
KW primary thrombocytopenia; cancer; adenocarcinoma; leukaemia; myeloma;
KW sarcoma; immune system disorder; acquired immunodeficiency syndrome;
KW AIDS; allergy; asthma; Crohn's disease; diabetes mellitus; gout;
KW glomerulonephritis; Goodpasture's syndrome; thyroiditis; pancreatitis;

KW hepatitis; multiple sclerosis; osteoporosis; Reiter's syndrome;
KW rheumatoid arthritis; neurological disorder; epilepsy; stroke; dementia;
KW Alzheimer's disease; Pick's disease; Huntington's disease; mood; anxiety;
KW Parkinson's disease; central nervous system disorder; mental disorder;
KW schizophrenic disorder; amnesia; Tourette's disorder; transgenic animal;
gene therapy.
OS Homo sapiens.
XX WO200257304-A2.
XX 25-JUL-2002.
XX 15-JAN-2002; 2002WO-US01340.
XX 16-JAN-2001; 2001US-261864P.
XX 16-JAN-2001; 2001US-261865P.
XX 16-JAN-2001; 2001US-261979P.
XX 16-JAN-2001; 2001US-261981P.
XX 17-JAN-2001; 2001US-262164P.
XX 17-JAN-2001; 2001US-262208P.
XX 17-JAN-2001; 2001US-263131P.
XX 19-JAN-2001; 2001US-262599P.
XX 19-JAN-2001; 2001US-262760P.
XX 19-JAN-2001; 2001US-263063P.
XX 19-JAN-2001; 2001US-263066P.
XX 19-JAN-2001; 2001US-263069P.
XX 19-JAN-2001; 2001US-263070P.
XX 19-JAN-2001; 2001US-263074P.
XX 19-JAN-2001; 2001US-263076P.
XX 19-JAN-2001; 2001US-263077P.
XX 19-JAN-2001; 2001US-263329P.
XX (INCY-) INCYTE GENOMICS INC.
XX Panzer SR, Lincoln SE, Altus CM, Dufour GE, Hillman JL, Jones AL;
PI Dam IC, Liu TF, Harris B, Flores V, Daffo A, Marwaha R, Chen AU;
PI Chang SC, Gerstin EH, Peralta CH, David MH, Lewis SA;
XX WPI; 2002-590716/63.
XX N-PSDB; ABS51309.
XX New purified secretory polypeptides and polynucleotides, useful in the
PT diagnosis, study, prevention or treatment of diseases associated with
PT decreased expression of functional secretory molecules, e.g. AIDS,
PT cancer or allergies -
XX Claim 27; Page 304-305; 340pp; English.
XX The invention describes an isolated polynucleotide a naturally occurring
CC polynucleotide sequence at least 90 % identical to it, a polynucleotide
CC complementary to it or an RNA equivalent of it. The purified secretory
CC polypeptides (SPTM) and polynucleotides are useful in the diagnosis,
CC study, prevention or treatment of diseases associated with decreased
CC expression of functional SPTM, e.g. actinic keratosis, arteriosclerosis,
CC bursitis, cirrhosis, hepatitis, polycythaemia vera, primary
CC thrombocytopenia, anaemia, psoriasis, cancers including adenocarcinoma,
CC leukaemia, myeloma or sarcoma, immune system disorder such as acquired
CC immunodeficiency syndrome (AIDS), allergies, asthma, Crohn's disease,
CC diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout,
CC Hashimoto's thyroiditis, hepatitis, multiple sclerosis, osteoporosis,
CC pancreatitis, Reiter's syndrome, autoimmune thyroiditis or rheumatoid
CC arthritis, neurological disorders such as epilepsy, stroke, Alzheimer's
CC disease, Pick's disease, Huntington's disease, dementia, Parkinson's
CC disease, other developmental disorder of the central nervous system,
CC mental disorder including mood, anxiety or schizophrenic disorder,
CC amnesia or Tourette's disorder. The polynucleotides may be used in
CC hybridisation and amplification technologies, e.g. in assessing gene
CC expression patterns, to develop a transcript image for a particular cell
CC or tissue, or to create transgenic animals to model human disease. This
CC is the amino acid sequence of a human secretory protein isolated in the
CC invention.


```
SQ Sequence 239 AA;
Alignment Scores:
Pred. No.: 7,99e-109 Length: 239
Score: 1237.00 Matches: 238
Percent Similarity: 99.58% Conservative: 0
Best Local Similarity: 99.58% Mismatches: 1
Query Match: 22.74% Indels: 0
DB: 23 Gaps: 0

US-10-001-857-42 (1-3096) x ABG69816 (1-239)
QY 475 GCGGCGGCGGAGGCGGCTCTTATTTCCGTGGTCCGACAGTGGTGGCGGCGGGTG 534
Db 1 AalalaGluAlaAaSerLeuPheProTrpSerGlyGlnCysValAlaAalargVal 20
QY 535 ACCACGGAGAGTAGGCAATATGTTATGAAGCTTCTGTAGATGATGAGATTCACGA 594
Db 21 ThrThrGlyGluValGlyIleMetValMetLysAlaSerValAspAspAspSerGly 40
QY 595 TGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAACAATACAACTGGGTGGACATACC 654
Db 41 TrpGluLeuSerMetProGluLysMetGluLysSerAsnThrAsnTrpValAspIleThr 60
QY 655 CAAGATTTTGAAGAGCTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCTA 714
Db 61 GlnAspPheGluGluAlaCysArgGluLeuLysLeuGlyGluLeuLeuHisAspLysLeu 80
QY 715 TTTGGTCTTTTGAAGCCATGCTCTCTATTGAAATGATGGATCCCAAGATGATGCTGC 774
Db 81 PheGlyLeuPheGluAlaMetSerAlaIleGluMetMetAspProLysMetAspAlaGly 100
QY 775 ATGATTGAAACCAAGTAAATCGAAAAAGTTCTCAATTTTGAACAGCTATCAAGGATGC 834
Db 101 MetIleGlyAsnGlnValAsnArgLysValLeuAsnPheGluGlnAlaIleLysAspGly 120
QY 835 ACTATTAAATTAAGATCTCACCTGCTGCACTGATGAGGATATGATGATACATGTTT 894
Db 121 ThrIleLysIleLysAspLeuThrLeuProGluLeuIleGlyIleMetAspThrCysPhe 140
QY 895 TGCTGTTTGATAACGTGTTAGAGGCCATTCACCTGGCACAGACAGATATTACGGCTT 954
Db 141 CysCysLeuIleThrTrpLeuGluGlyHisSerLeuAlaGlnThrValPheThrCysLeu 160
QY 955 TACATTCTATATCCAGATCTTATAGAACTCTGCTATGAAGCTTTTGTCTGGGATC 1014
Db 161 TyrIleHisAsnProAspPheIleGluAspProAlaMetLysAlaPheAlaLeuGlyIle 180
QY 1015 TTGAAATCTGTGACATTGCAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGAA 1074
Db 181 LeuLysIleCysAspIleAlaArgGluLysValAsnLysAlaAlaValPheGluGlu 200
QY 1075 GATTTTCATCAATGACTATGATGATTTAAATGGCTAACTGTCAGATGTCAGATCTTCGATT 1134
Db 201 AspPheGlnSerMetThrTyrGlyPheLysMetAlaAsnSerValThrAspLeuArgVal 220
QY 1135 ACAGGCATGCTAAAGATGTGAGGATGACATGCACATGCACAAAGAGTAAAGACTCTGA 1191
Db 221 ThrGlyMetLeuLysAspValGlyAspAspMetGlnArgArgValLysSerThrArg 239

RESULT 5
AAU32414
ID AAU32414 standard; Protein; 108 AA.
XX
XX
AC AAU32414;
XX
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2905.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
```

Claim 20: SEQ ID No 44508; 103pp; English.

CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG0377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 156 AA;

Alignment Scores:
Pred. No.: 1.61e-05 Length: 156
Score: 153.50 Matches: 52
Percent Similarity: 34.66% Conservative: 9
Best Local Similarity: 29.55% Mismatches: 56
Query Match: 2.79% Indels: 59
DB: 22 Gaps: 7

US-10-001-857-42 (1-3096) x ABG14149 (1-156)

QY 556 TTATGCTACTTCTCCGCTGCTACCGCGCGCCGACGCACTGTCCGACCGAAATA 497
Db 1 LeuCysSerLeuSerLeuProProPro----- 12
QY 496 ACAGACGGCTCGCGCGCGCCGCCCTCTCAGCCAGCCGACGCGT 437
Db 13 ---ProProProProProProPro----- 22
QY 436 GCAGCATGCTATGCGCGGACGCGACCCGCCCTCTCCCTTACTCCCGGGACT 377
Db 23 -----ProProProProProProProProProValProLeuProPro----- 37
QY 376 TCGCGTCTCCCGGACACACCCCTCCCTCGCCACCGCTCGGTC-----GGCGGA 326
Db 38 -----SerProArgSerProProValSerProProProHisSerPheGlnGlyArg 55
QY 325 TGGCGCGCGCTCTCTTCGCGAGATGAGACGATCGGAGACCGGAAGTTATCTTT 266
Db 56 SerProProSerGluLeuGlySerSerAlaAluProTrpLeuArgProGlyThrTrpVal 75
QY 265 TTTCACCT-----CTGTGGAGTCTCGGTAGTGGCGGTTCGCGAAAGAGCG 215
Db 76 ***ProProProLeuThrPheSerGlnGlnGluAlaSerSerAlaGlnLys***Asn 95
QY 214 GAGCGCGAGTCTCA---GAGCGCGCCGCTCTCGCTGCGCTTCTGGAAATTTCTTAGAA 158
Db 96 CysProGlyGluSerProGlnProValProValThrValPro----- 109
QY 157 AGGTGACGGAAGCGGAAGAACCGGTGAGGTTTGGTAGACCTTAGTTGCGAGC 98
Db 109 ----- 109
QY 97 ACGAACGAGACCAAGCTCCGAGAGCAACAGCGTGTCTACTGTGCT 50
Db 110 ---GluArgGlnProAspAlaAlaProAsnHisSerLeuLeuLeuPro 124

RESULT 8

ABB91504
ID ABB91504 standard; Protein; 708 AA.

AC ABB91504;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 715.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX WO200210210-A2.

PD 07-FEB-2002.

XX

PF 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidler M;

DR WPI; 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -

XX Claim 5; SEQ ID NO 715; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.

SQ Sequence 708 AA;

Alignment Scores:
Pred. No.: 4.11e-05 Length: 708
Score: 153.00 Matches: 59
Percent Similarity: 39.38% Conservative: 17
Best Local Similarity: 30.57% Mismatches: 71
Query Match: 2.78% Indels: 46
DB: 23 Gaps: 8

US-10-001-857-42 (1-3096) x ABB91504 (1-708)

QY 508 ATACTGAGTCCCATCTGATCTCATCTACAGAGTTCATTAACCATATGCT 549
Db 101 ValileProSerProProSerAlaSerProProAlaLeuValProProLeuPro 120
QY 548 ACTTCTCC-----GTGCTACCCGCGCGCCACG 519
Db 121 SerSerProProProAlaSerValProProArgPro-SerProSerProI 140
QY 518 CACTGTCCGAGC-----ACGGAATAACACGCGCTCGGCGCGC-- 476
Db 140 eLeuValArgSerProProSerValArgProIleGlnSerProProProSe 160
QY 475 -----CGGCGCGCTCCCTCTCAGCCGACGCCGACGCGGTGC 435
Db 160 rAspArgProThrGlnSerProProSerProSerProSerProSerGluArgPr 180
QY 434 ACGCATG-----CGTATGCCCGGACGCGCGACCGCCCGCTC 396
Db 180 oThrGlnSerProSerProSerGluArgProThrGlnSerProProProSe 200
QY 395 TCCCTTACTCCGCGGACTTCGCGCTCCCGGCGCACACCCCTCCCTCGCCACGCTCC 336
Db 200 rProSerProSerAspArgProSerGlnSerPro-----ProProProGl 218
QY 335 GTTCGCGCGATGGCGCGCTTCTTCTTCGCGAGATTAGACGATCGGACCGGA 276
Db 218 uAspThrLysProGlnProProArg-----ArgSerProAsnSerProPr 233
QY 275 AGTTATCTTTTTCACCTCTGTGGAGTCTGGGTAGTGGCGGTTCGCGAAGAGGC 216
Db 233 oProThrPheSerSerPro-ProArgSerProProGluLeuValProGlySerAsn 253
QY 215 GGAGCCGCGAGTCTCAGAGCCCGCGCTGCGCTGCGCTCTGGGAATTTCTTAGAAG 156

